BMC Bioinformatics

RESEARCH

Open Access

Cancer survival prediction by learning comprehensive deep feature representation for multiple types of genetic data

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] School of Computer, Guangdong University of Petrochemical Technology, Maoming, China
[3] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

Abstract

Background: Cancer is one of the leading death causes around the world. Accurate prediction of its survival time is significant, which can help clinicians make appropriate therapeutic schemes. Cancer data can be characterized by varied molecular features, clinical behaviors and morphological appearances. However, the cancer heterogeneity problem usually makes patient samples with different risks (i.e., short and long survival time) inseparable, thereby causing unsatisfactory prediction results. Clinical studies have shown that genetic data tends to contain more molecular biomarkers associated with cancer, and hence integrating multi-type genetic data may be a feasible way to deal with cancer heterogeneity. Although multi-type gene data have been used in the existing work, how to learn more effective features for cancer survival prediction has not been well studied.

Results: To this end, we propose a deep learning approach to reduce the negative impact of cancer heterogeneity and improve the cancer survival prediction effect. It represents each type of genetic data as the shared and specific features, which can capture the consensus and complementary information among all types of data. We collect mRNA expression, DNA methylation and microRNA expression data for four cancers to conduct experiments.

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

Availability and implementation: https://github.com/githyr/ComprehensiveSurvival.

Keywords: Cancer survival prediction, Shared information, Specific information, Comprehensive representation

Introduction

As the morbidity and mortality rates gradually rise, cancer is becoming the main death cause in the global [1–3]. According to the global cancer report, additional 14.10 million cancer cases occurred with death cases 8.20 million in 2012. The number of new cancer cases and death cases reached 18.1 million (9.5 million men and 8.6 million women) and

9.6 million respectively in 2018 [4]. Meanwhile, cancers have become common among young people. Therefore, it is significant to accurately predict the survival time, which can help the clinicians make proper therapeutic guidance to improve the survival rate and living quality of cancer patients [5, 6].

Cancer survival prediction has been an interesting and challenging issue in cancer research over the past few decades [7–9]. The heterogenous disease, cancer, can be characterized by varied molecular features, clinical behaviors, morphological appearances and reactions to therapies. This leads that the genes and phenotypes of cells in the same pattern and stage are also different, which results in a big challenge for cancer survival prediction [10–12]. Figure 1 visualizes the embedding feature spaces of DNA methylation, mRNA expression and microRNA expression for the glioblastoma multiforme (GBM) dataset by reducing the dimensionality of original features. Specifically, we use a commonly-used visualization method t-SNE (t-distributed stochastic neighbor embedding) [13] to display the low-dimensional feature space of genetic data. t-SNE adopts the nonlinear dimensionality reduction technique and can preserve the local and global distribution structures of the dataset. As shown in this figure, patient samples with different survival times are mixed together and difficult to be distinguished in the embedding feature spaces of three types of genetic data, which further verifies the difficulty of cancer survival prediction.

Survival analysis is usually accomplished using heterogeneous data sources including low-dimensional clinical data (age, sex, cancer grade detail, body fat rate, etc.) [14], pathological images [15–18], and multi-type gene data [19]. For example, Chen et al. proposed an interpretable strategy for end-to-end multimodal fusion of histology image and genomic (mutations, CNV, RNA-Seq) features [20]. Cheerla et al. designed an unsupervised encoder to integrate four data modalities (gene expression data, miRNA data, clinical data and whole slide image) into a single feature vector for each patient [21]. Vale-Silva et al. utilized clinical, imaging, and different high-dimensional omics data modalities to conduct cancer survival prediction [22]. Compared with single source data, multi-source heterogeneous data describes the cancer from different perspectives, which can capture a more comprehensive understanding of the cancer [23, 24]. Multi-source heterogeneous data can be regarded as multi-modal data, which contains not only large consensus information but also abundant complementary information [25]. From information perspective, consensus information indicates that each modality contains information that shared by all modalities (inter-modal shared information);

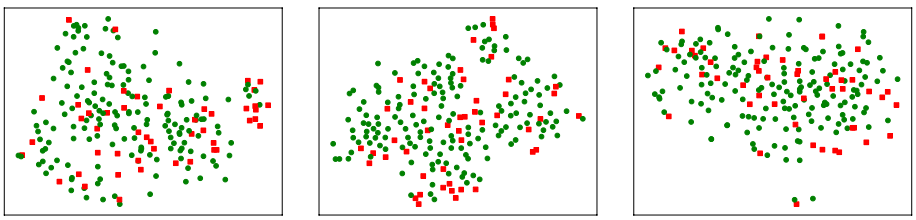


Fig. 1 Visualizations of embedding feature spaces of DNA methylation, mRNA expression and microRNA expression for GBM. The red dots point short time survivors (< 2 years)and the blue dots represent long time survivors (> 2 years)

complementary information instructs that each modality also contains information that is unique to itself (intra-modal specific information) [26].

Existing data integration based cancer survival prediction methods can be mainly categorized into three paradigms:

(i) fusion methods [27–30] based on concatenation integrate multiple types of data directly. This seems unreasonable since the concatenation of heterogeneous data sources neglects the inter-modal discriminant information. In addition, this strategy would cause very high dimensional feature vectors, which is adverse for feature learning [31].

(ii) fusion methods [21, 32–34] only learn consensus information. This strategy only exploits the consensus information of heterogeneous data but ignores the diversity of heterogeneous data, which is adverse for exploiting comprehensive information of cancer. While, for heterogenous disease, making full use of the complementarity between different types of data is conducive to a comprehensive understanding of the disease.

(iii) works [25] and [35] utilize the similarity network fusion to integrate multiple types of data. They learn consensus and complementary information based on the relations between patient samples. However, they ignore fine-grained feature representation information, especially for gene sequences with thousands of dimensions.

In general, although multi-type gene data have been used in the existing work, how to learn more effective features for cancer survival prediction and explain them at the feature level has not been well studied. Besides, deep learning technique has been proved to have strong feature representation and classification ability in various tasks. In this paper, we intend to utilize deep learning to obtain more effective feature representations of multi-type genetic data and achieve better performance of cancer survival prediction at the feature level, which can reduce the negative impact of data heterogeneity. Also, we want to explain the functions of deep features from the aspect of extracting consensus and complementary information. Figure 2 shows a demo of the proposed deep learning for cancer survival prediction. These consensus and complementary representations are exploited to capture comprehensive survival information of cancer patients; e.g., consensus representation is exploited to capture modality-invariant survival information; the specific representations of mRNA expression, DNA methylation and miRNA expression are exploited to capture the modality-specific survival information.

The main contributions of this study are summarized as follows:

(1) This study focuses on the problem of data heterogeneity in cancer survival prediction and proposes a deep learning approach to integrate multi-type genetic data effectively. As shown in Fig. 2, by sufficiently integrating multi-type genetic data, survivors with different times (i.e., short and long times) can be well separated in the feature space built by our approach, which means that the negative impact of data heterogeneity on cancer survival prediction can be alleviated significantly.

(2) In the proposed deep learning approach, it represents each type of genetic data as the shared and specific features, which can capture the consensus and complementary information among all types of data. Then, we fuse the shared and specific fea-

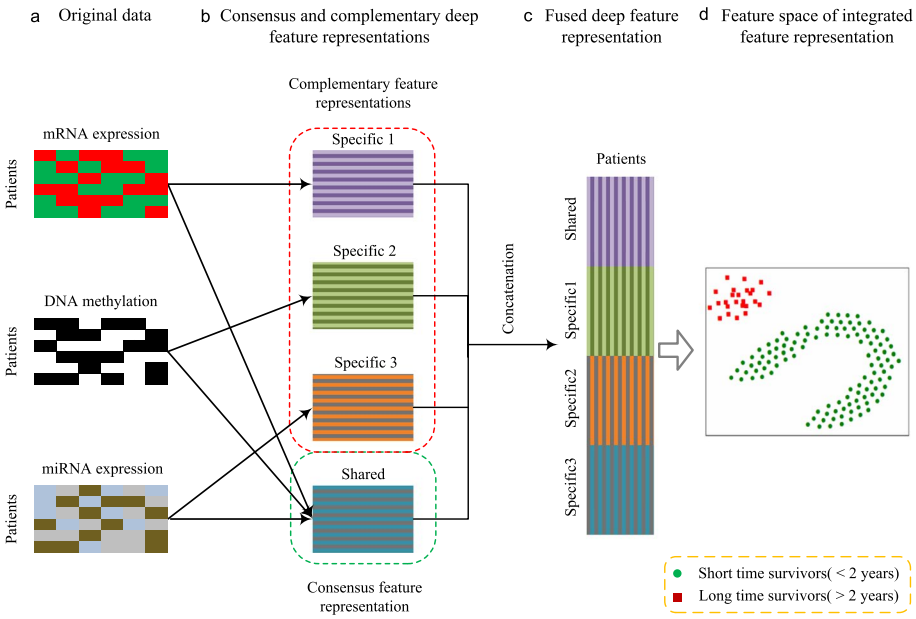


Fig. 2 A demo for deep learning for cancer survival prediction. **a** Example representations of mRNA expression, DNA methylation and miRNA expression datasets for the same cohort of patients. **b** Learning deep feature representation from the perspective of shared and specific information. **c** Fused deep feature representation by concatenation for cancer survival prediction. **d** The visualization of embedding feature space of integrated feature representation

tures of each type of data by concatenation, and employ the fused features for cancer survival prediction as shown in Fig. 2. To strengthen the representation ability of deep features, we layer-by-layer impose an Euclidean distance constraint on the shared feature learning network, as well as impose an orthogonal constraint on the specific feature learning network.

(3) We conduct extensive experiments on glioblastoma multiforme (GBM), kidney renal clear cell carcinoma (KRCCC), lung squamous cell carcinoma (LSCC) and breast invasive carcinoma (BIC) datasets. Experimental results show that our approach can achieve higher prediction performance than competing methods. This demonstrates that our approach significantly improves the performance of cancer survival prediction and is helpful for clinicians to make proper therapeutic guidance for cancer patients.

Proposed methods

Figure 3 shows the proposed deep learning network to achieve the shared and specific feature representation for cancer survival prediction. First, it maps the original feature dimensions of all data types to the same dimension through a fully connected neural network. Secondly, it builds a multi-stream deep shared network with parameters shared for all data types to learn the consensus information, as well as a deep specific network for each data type to learn the complementary information. At the same time, an Euclidean distance constraint is used to enhance the learning of consensus information and an orthogonal constraint is used to enhance the learning of complementary information.

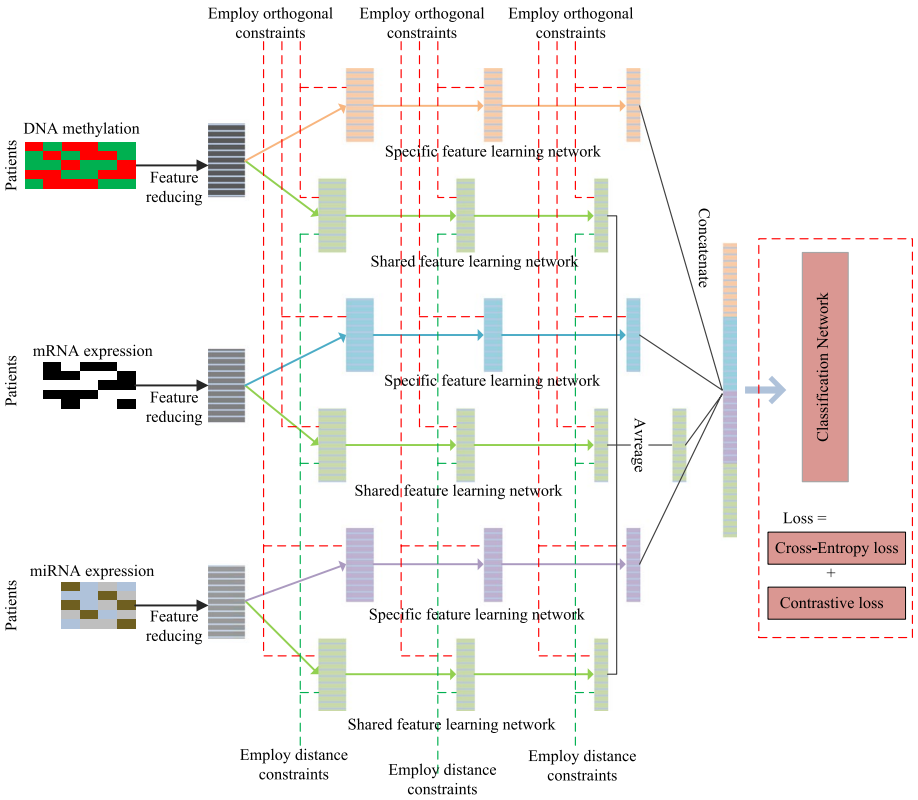


Fig. 3 The architecture of our proposed deep learning approach

Finally, to improve the separability of data, we introduce the contrastive loss to pull samples from the same class closer and push samples from different classes farther.

Feature mapping

The data integration strategy designed in this paper can learn the consensus and complementary information only when the feature dimensions of these data types are consistent. Therefore, it is necessary to map the features of all genetic data types to the same dimension. A common approach is to adopt the Max-Relevance and Min-Redundancy (mRMR) feature selection algorithm for dimension reduction [27, 36, 37], which ignores the interaction between gene sites in sequence. In this paper, we design a three-layer fully connected neural network for feature mapping. Considering that the dimension increase operation will introduce noise, we employ dimension reduction operation to get the same feature dimension for all data types. In addition, the data dimension for miRAN is relatively low (329 dimension for KRCCC to 534 dimension for GBM), which is not suitable for further dimensionality reduction. Therefore, we use the dimension of miRNA as the last mapping dimension. Table 1 shows the detailed dimensionality values for the feature mapping process.

Shared and specific deep feature learning

Let $X = \{x_i \in \mathbb{R}^q\}_{i=1}^N$ be a set of N samples, where q represents dimension of each sample. Moreover, let $X_K = \left\{x_{k,i} \in \mathbb{R}^{q^k}\right\}_{i=1}^N$ denote the feature set of X in data type k, where $x_{k,i}$ is

Table 1 Feature mapping process of three data types for four cancer datasets

Datasets	Modality	Dimensionality	The dimension of ANN	Last mapping dimension
GBM	mRNA	12042	12042→4096→534	534
	miRNA	534	534→534→534	534
	DNA	1305	1305→768→534	534
KRCCC	mRNA	17899	17899→4096→329	329
	miRNA	329	329→329→329	329
	DNA	24960	24960→4096→329	329
LSCC	mRNA	12042	12042→4096→352	352
	miRNA	352	352→352→352	352
	DNA	23074	23074→4096→352	352
BIC	mRNA	17814	17814→4096→354	354
	miRNA	354	354→354→354	354
	DNA	23094	23094→4096→354	354

the k-th representation of the x_i and q^k is the dimension of x_i. Here, $k = 1, 2, \ldots, K$, where K denotes the total number of data types. Generally, the $k-$th representation $\mathrm{x}_{k,i}$ and the $l-$th representation $\mathrm{x}_{l,i}$ of the x_i, $k \neq l$, are different, because they are usually from different spaces. Therefore, directly concatenating them may not be physically meaningful, and cannot well utilize the complementary property.

Considering the fact that each data type represents the same object from different point of view, different data types not only contain the specific information but also share common information. For $\mathrm{x}_{k,i}$, we employ the shared feature learning network to project it to get the consensus information by $h^c_{k,i} = W^c_k \mathrm{x}_{k,i}$, where $W^c \in \mathbb{R}^{r^c \times q_k}$, and employ the specific feature learning network to project it to get the complementary information by $h^s_{k,i} = W^s_k \mathrm{x}_{k,i}$, where $W^s_k \in \mathbb{R}^{r^s_k \times q_k}$. The learned feature representation of $\mathrm{x}_{k,i}$ can be written as:

$$h_{k,i} = \begin{pmatrix} h^s_{k,i} \\ h^c_{k,i} \end{pmatrix} = \begin{pmatrix} W^s_k \\ W^c_k \end{pmatrix} \mathrm{x}_{k,i}. \tag{1}$$

Therefore, the final representation with multiple data types can be denoted as:

$$h_i = \left[{h^s_{1,i}}^T, {h^s_{2,i}}^T, \ldots, {h^s_{K,i}}^T, {h^c_{1,i}}^T, {h^c_{2,i}}^T, \ldots, {h^c_{K,i}}^T \right]^T. \tag{2}$$

Since the shared information from different data types is almost the same, it is unnecessary to include all of them in the final representation. Instead, we use the average value:

$$h^c_i = W^c \mathrm{x}_i \triangleq \frac{1}{K} \sum_{k=1}^{K} W^c_k \mathrm{x}_{k,i}. \tag{3}$$

Finally, the resulting representation of x_i can be written as:

$$h_i = \left[{h^s_{1,i}}^T, {h^s_{2,i}}^T, \ldots, {h^s_{K,i}}^T, {h^c_i}^T \right]^T. \tag{4}$$

Layer-by-layer Euclidean distance and orthogonality constraints

We impose the orthogonality constraint between each layer of shared and specific feature learning networks to separate shared and specific information, as well as prevent them from contaminating each other. Furthermore, we impose the Euclidean distance constraint between each layer of multi-stream shared feature learning networks to ensure the similarity of consensus information. Details are described as follows:

Let $H_k^c(m)$ and $H_k^s(m)$ be the outputs of shared and specific networks from layer m. Orthogonality loss between $H_k^c(m)$ and $H_k^s(m)$ is defined as:

$$L_{diff} = \left\| H_k^c(m)^{\mathrm{T}} H_k^s(m) \right\|_F^2, \tag{5}$$

where $\|\cdot\|_F^2$ is the squared Frobenius norm.

Let $H_k^c(m)$ and $H_l^c(m)$ be the outputs of the same layer for data type k and l in shared feature learning network, respectively. Euclidean distance loss between $H_k^c(m)$ and $H_l^c(m)$ is defined as:

$$L_{diss} = \frac{1}{2N} \sum_{n=1}^{N} d_n^2, \tag{6}$$

where $d_n = \left\| h_{k,n}^c - h_{l,n}^c \right\|^2$, and $h_{k,n}^c$ and $h_{l,n}^c$ are shared representation of sample x_n in data type k and l, respectively.

Classification

After integrating multiple genetic data types into a unified representation, we classify them with a multilayer network. Cross-Entropy loss is used for classification.

To improve the separability of data, contrastive loss is implemented. Specifically, for a pair of samples x_i and x_j, we use h_i and h_j to represent their features extracted by the feature learning network, respectively. The distance between them is computed as:

$$d(x_i, x_j) = \left\| h_i - h_j \right\|_2. \tag{7}$$

Contrastive loss between h_i and h_j is defined as:

$$L_{con} = \frac{1}{2N} \sum_{n=1}^{N} \left[y_n d_n^2 + (1 - y_n) \max^2(Margin - d_n, 0) \right], \tag{8}$$

where d_n is the distance of the n^{th} paired samples, *Margin* is a threshold, and y_n denotes whether the paired samples are from the same class. If they are from the same class, $y_n = 1$, otherwise, $y_n = 0$.

Experiments

Datasets

Four cancer datasets 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 approach. For each dataset, we collect three types

of genetic 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 [25]. It downloads these data from the 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 a certain 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, otherwise, the k-nearest interpolation is used for complementing this gene; (iii) the z-score transformation is used for normalizing the data samples.

Figure 4 illustrates the survival time distribution for four cancer datasets, from which we can observe that the survival time for GBM, KRCCC, LSCC and BIC ranges 0–118 months, 0–113 months, 0–125 months and 0–192 months, respectively. The median survival for GBM, KRCCC, LSCC and BIC is 14 months, 45 months, 19 months and 26 months, respectively. Combined with the survival time distribution and median survival of each cancer, 2-year, 4-year, 2-year and 3-year are taken as thresholds to divide two types of patients with four cancer types. Table 2 shows the data properties of four datasets. For classification, the short term patients are labeled as 0 and long term patients are labeled as 1. The initial feature dimensions of three types of genetic expressions in all datasets are significantly different.

Evaluation

To evaluate our proposed method, we adopt ten-fold cross validation in our experiments. Specifically, we randomly divide long time survivors and short time survivors into ten subsets, respectively. For each round of training, each subset of long time survivors combined with each subset of short time survivors will be used as a validation set, another seven subsets of long time survivors combined with seven subsets of short time survivors are used as training set, the last two subsets of long time survivors combined

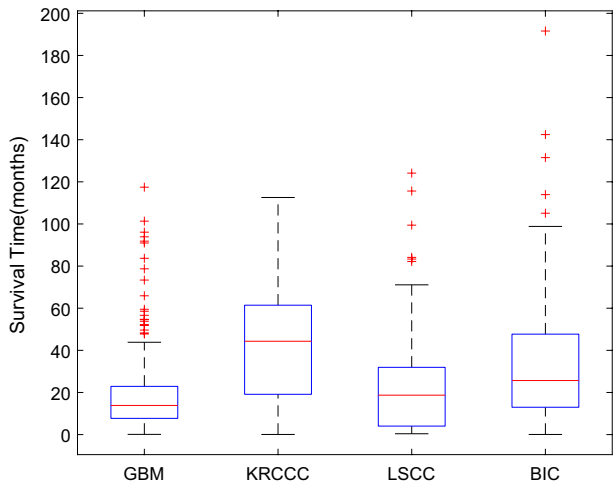


Fig. 4 Survival time distribution in four cancer datasets as represented by box plots (center red line represents median, lower and upper quartiles and whiskers capture max and min values of the survival time in each cancer)

Table 2 Data properties of four cancer datasets

Datasets	Instance	Cut-off (years)	Short/long time survivors	Modality	Dimensionality
GBM	215	2	166/49	mRNA	12042
				miRNA	534
				DNA	1305
KRCCC	122	4	67/55	mRNA	17899
				miRNA	329
				DNA	24960
LSCC	106	2	66/40	mRNA	12042
				miRNA	352
				DNA	23074
BIC	105	3	67/38	mRNA	17814
				miRNA	354
				DNA	23094

with the last two subsets of short time survivors are used as testing set. The prediction score is the average of the output of ten rounds. In this paper, we use five metrics including Accuracy (*Acc*), Recall, Precision(*Pre*), ROC curve and AUC (area under the ROC curve) to measure model performance. These metrics are defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{9}$$

$$Recall = \frac{TP}{TP + FN}, \tag{10}$$

$$Pre = \frac{TP}{TP + FP}, \tag{11}$$

where true positive (TP) represents the number of cases correctly identified as short-survival, false positive (FP) represents the number of cases incorrectly identified as short-survival, true negative (TN) represents the number of cases correctly identified as long-survival, and false negative (FN) represents the number of cases incorrectly identified as long-survival.

Hyper-parameter selection

The designed cancer survival prediction model consists of three modules: features mapping network, shared and specific representation learning network and classification network. Specifically, the features mapping network adopts a three-layer fully connected network, and the size of each layer is shown in Table 1. We build the shared and specific representation learning network with two hidden fully connected layers of sizes 256 and 128, and an output layer of size 32. Each layer uses the ReLU activation function. The classification network adopts a three-layer fully connected network, in which the sizes of hidden and output layers are 32 and 2, respectively.

To avoid overfitting, we do not perform a separate hyper-parameter search for each cancer dataset. Instead, we search the hyper-parameters on GBM dataset and apply the

selected parameters for other datasets. The hyper-parameter margin is searched on the grid [1, 2, 3, 4, 5]. We perform grid search based on the grid [0.0001, 0.0003, 0.0005, 0.0007, 0.0009, 0.001] to determine the learning rate of Adam optimizer. Batch size for training set is searched from [20, 30, 40, 50]. Specifically, we conduct a series of tests on the validation set where in each experiment we vary one of the three hyper-parameters from the chosen value by tuning it up or down by one grid, obtaining 15 sets of varied hyper-parameters. For each set of varied hyper-parameters, 10-fold cross-validation is conducted.

The final chosen hyper-parameters are shown in Table 3.

Experimental results

We compare our approach with three state-of-the-art cancer survival prediction methods:

- Similarity network fusion (SNF) for aggregating data types on a genomic scale [25];
- Integrating multiple genomic data and clinical data based on graph convolutional network (GCGCN) for cancer survival prediction [35];
- Multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data (MDNNMD) [28].
- Multi-modal advanced deep learning architectures for breast cancer survival prediction (SiGaAtCNNs) [30];
- Cross-aligned multimodal representation learning for cancer survival prediction (CAMR) [38];
- Integrating multi-omics data by learning modality invariant representations for improved prediction of overall survival of cancer (LMIR) [39].

A brief introduction of these survival analysis methods is summarized in Table 4. The predictive results of all competing methods are reported in Figs. 5 and 6. Figure 5 shows the comparison results of all evaluation metrics including accuracy, precision and the area under curve (AUC) on four datasets. Figure 6 presents the receiver operating characteristic (ROC) curves of all competing methods on four datasets.

From these results, we can conclude that the overall performance of our method is much higher than those of three compared methods. This indicates that methods considering consensus and complementary information are better than that simply concatenating features.

In order to further investigate the effectiveness of learned feature representations by our approach, i.e., the final fusion representation by concatenating all specific representations and the shared representation, we employ the t-SNE to embed the samples

Table 3 The selected hyper-parameters for prediction model

Hyper-parameter	Value
Margin	2.0
Learning rate	0.0003
Batch size	30

Table 4 Brief introduction of our approach and three compared methods

Methods	Deep learning	Data-integration strategies		
		Simple concatenation at feature level	Learning consensus and complementary information at sample level	Consensus and complementary information learning at feature level
SNF	No	No	Yes	No
GCGCN	Yes	No	Yes	No
MDNNMD	Yes	Yes	No	No
SiGaAtCNNs	Yes	Yes	No	No
CAMR	Yes	No	No	Yes
LMIR	Yes	Yes	No	No
Ours	Yes	No	No	Yes

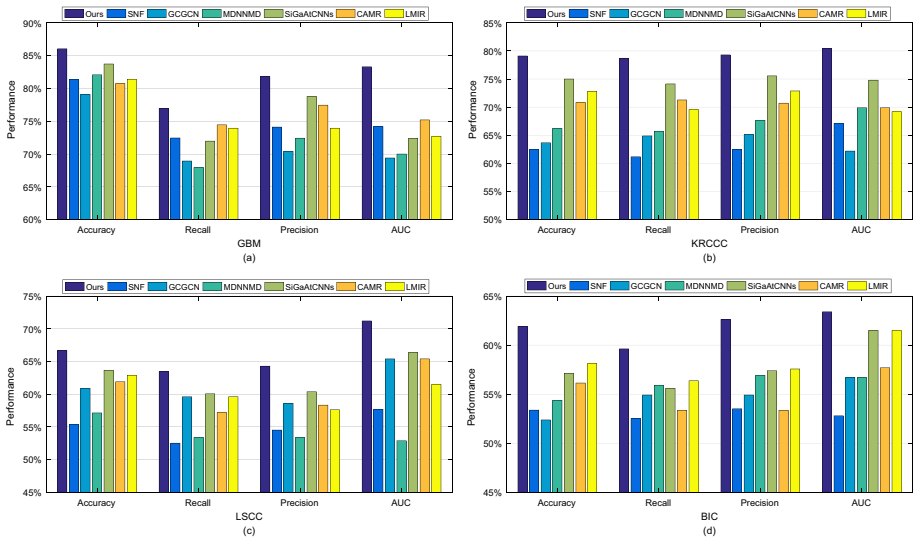


Fig. 5 Prediction performance for four cancers survival prediction, comparing SNF, GCGCN, MDNNMD, SiGaAtCNNs, CAMR, LMIR and ours

into the two-dimensional space for visualization. Figure 7 illustrates the distribution of original training samples and the distribution of learned feature representations on four cancer datasets. From the figure, we can observe that (1) t-SNE produces visually interpretable results by converting vector similarities into joint probabilities, generating visually distinct clusters that represent patterns in the data. (2) the samples with different survival stages are mixed together and not well separated in the original feature space. (3) With the learned shared features, specific features of mRNA, specific features of miRNA, and specific features of DNA, patients with the same survival stage tend to be clustered. (4) With the final learned integrated features, the samples from different survival stages can be intuitively separated into two disjoint clusters, which indicates the better separability of integrated feature representations.

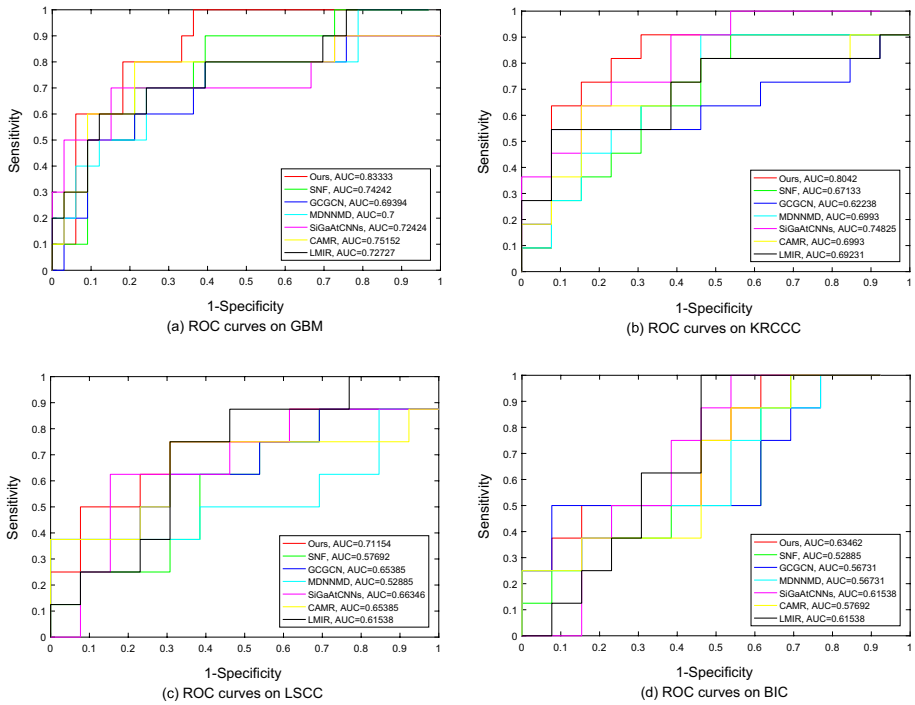


Fig. 6 ROC curves for four cancers survival prediction, comparing SNF, GCGCN, MDNNMD, SiGaAtCNNs, CAMR, LMIR and ours

Survival analysis

Survival analysis expresses a statistical method considering both results and survival time. Figure 8 shows the confusion matrixes of test sets on four cancers. The Kaplan-Meier (KM) survival curves are drawn in Fig. 9, and their P values are calculated according to the curves. For GBM, KRCCC and LSCC, there are significant differences between high-risk and low-risk patients (p values are 8.70×10^{-5}, 1.8×10^{-4}, 3.23×10^{-4}, respectively), while for BIC, the difference is not significant ($p = 0.471$). The p values for GBM, LSCC, KRCCC and BIC rise significantly when the censored data ratio rises from 0.077 for GBM to 0.875 for BIC. The reason is that the model can hardly learn well by primarily using the censored data.

Effect of layer-by-layer constraints for strengthening feature representation ability

To investigate the effect of layer-by-layer constraints in our approach, we construct the compared backbone by *i*mposing *c*onstraints only on the *l*ast *l*ayer of deep learning network and denote it as ICLL. Figure 10 reports the comparison results of ICLL versus ours. Overall, our approach performs better than ICLL on all datasets in terms of accuracy, precision and AUC. The average performance improvements are 5.00%, 4.75%, 2.75% and 7.50% on GBM, KRCCC, LSCC and BIC datasets respectively, which indicates the effectiveness of imposing distance and orthogonal constraints layer-by-layer.

There are two reasons that the proposed approach is superior to ICLL that only imposes the constraints on the last layer of deep learning network: (i) layer-by-layer imposing constraints learns the shared and specific features multiple times, which can

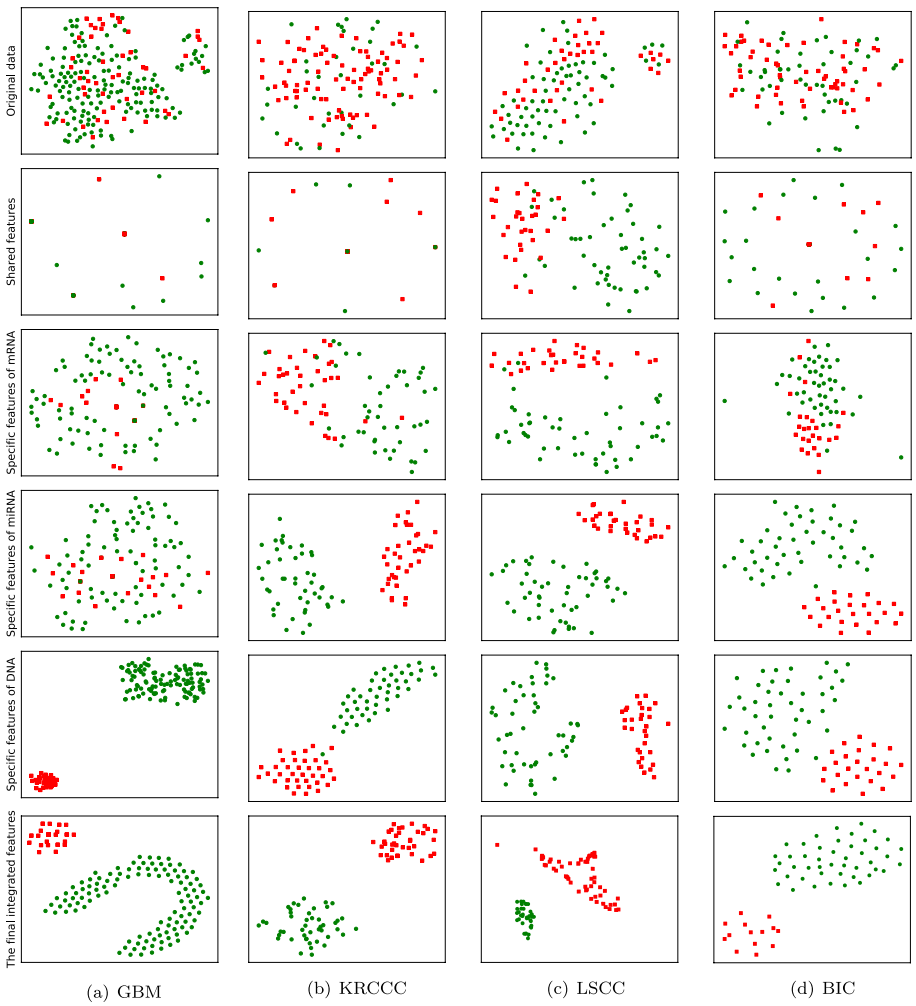
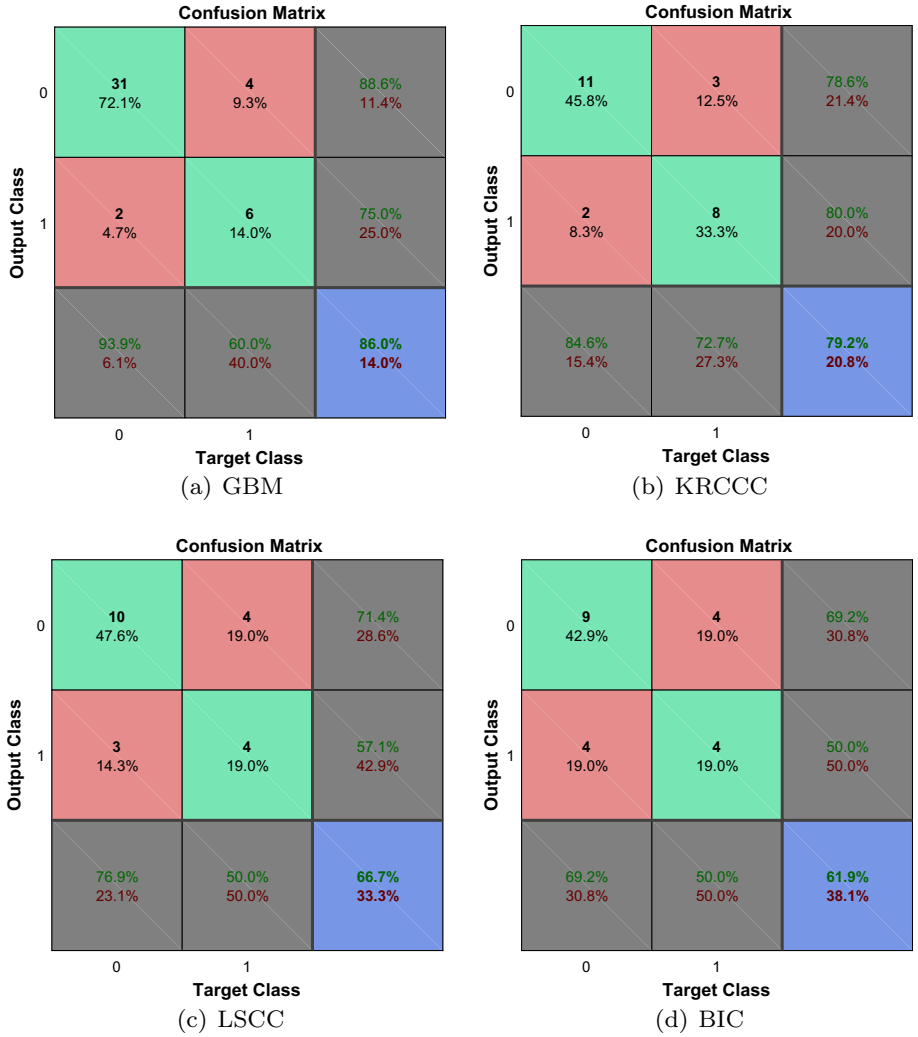


Fig. 7 T-SNE visualization of data on each dataset. The top plots in **a**–**d** present the distribution of original samples with concatenated features. The middle four plots in **a**–**d** show the distribution of shared features, specific features of mRNA, specific features of miRNA, and specific features of DNA, respectively. The bottom plots in **a**–**d** show the distribution of samples with learned features by our approach

obtain better consensus and complementary feature representation than ICLL that learns shared and specific features only one time; (ii) layer-by-layer constraints are employed on each layer of deep learning networks, which can avoid learning networks falling into local optimal solution and can learn robust representations.

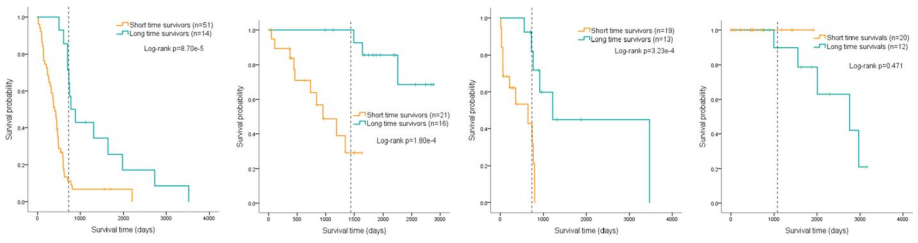
Conclusion

Accurate prediction of survival time of cancers is significant, which can help clinicians make appropriate therapeutic schemes. State-of-the-art works show that integrating multi-type genetic data may be an effective way to deal with data heterogeneity, but they cannot provide a rational and feature representation for multi-type genetic data. To this end, we propose a deep learning approach which can learn the consensus and complementary information between multi-type genetic data at the feature level. It explicitly represents each type of genetic data as the shared and specific features to



(a) GBM (b) KRCCC (c) LSCC (d) BIC

Fig. 8 Confusion matrixes of test sets on four cancers



(a) GBM (b) KRCCC (c) LSCC (d) BIC

Fig. 9 Kaplan–Meier curves of four cancers prognosis prediction. The dotted line in KM curve represents the median cut-off of two survivors

strengthen the interpretability. Sufficient experiments verify that the our approach can significantly improve the cancer survival prediction performance as compared with related works. In summary, our work provides an effective deep learning method to overcome data heterogeneity in cancer survival prediction.

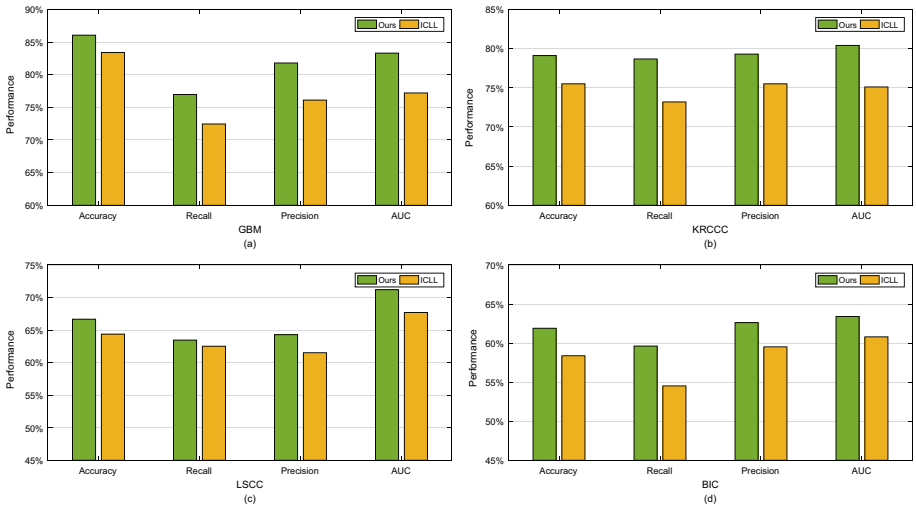


Fig. 10 Prediction performance comparison between ICLL and ours

Acknowledgements
Not applicable.

Author contributions
YH: conceptualization, methodology, writing- original draft preparation. X-YJ: writing- reviewing and editing, supervision, data curation. QS: visualization, investigation, software, validation.

Funding
This work was supported by the NSFC Project under Grant Nos. 62176069 and 61933013, the Innovation Group of Guangdong Education Department under Grant No. 2020KCXTD014, the Innovation Group of Guangdong Education Department under Grant No.2020KCXTD014.

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.

Received: 7 February 2023 Accepted: 19 June 2023
Published online: 28 June 2023

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics. CA Cancer J Clin. 2015;65(2):87–108.
2. 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.
3. 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.
4. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
5. 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.
6. Wang Y, Wang D, Ye X, Wang Y, Yin Y, Jin Y. A tree ensemble-based two-stage model for advanced-stage colorectal cancer survival prediction. Inf Sci. 2019;474:106–24.
7. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015;13(C):8–17.

8. Travers C, Zhu X, Garmire LX, Florian M. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. PLoS Comput Biol. 2018;14(4):1006076.
9. Luck M, Sylvain T, Cardinal H, Lodi A, Bengio Y. Deep learning for patient-specific kidney graft survival analysis. Preprint at 1705.10245 (2017)
10. Zhang H, Zheng Y, Hou L, Zheng C, Liu L. Mediation analysis for survival data with high-dimensional mediators. Bioinformatics. 2021;37(21):3815–21.
11. 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.
12. Cui L, Li H, Hui W, Chen S, Yang L, Kang Y, Bo Q, Feng J. A deep learning-based framework for lung cancer survival analysis with biomarker interpretation. BMC Bioinform. 2020;21(1):112.
13. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9(11):2579–605.
14. Louis DN, Perry A, Reifenberger G, Deimling AV, 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.
15. Shao W, Wang T, Huang Z, Han Z, Zhang J, Huang K. Weakly supervised deep ordinal cox model for survival prediction from whole-slide pathological images. IEEE Trans Med Imaging. 2021;40(12):3739–47.
16. Zhang L, Dong D, Liu Z, Zhou J, Tian J. Joint multi-task learning for survival prediction of gastric cancer patients using CT images. In: International symposium on biomedical imaging; 2021. p. 895–8.
17. Agarwal S, Abaker MEO, Daescu O. Survival prediction based on histopathology imaging and clinical data: a novel, whole slide CNN approach. In: Medical image computing and computer assisted intervention; 2021. p. 762–71.
18. Fan L, Sowmya A, Meijering E, Song Y. Learning visual features by colorization for slide-consistent survival prediction from whole slide images. In: Medical image computing and computer assisted intervention; 2021. p. 592–601.
19. Qiu YL, Zheng H, Devos A, Selby H, Gevaert O. A meta-learning approach for genomic survival analysis. Nat Commun. 2020;11:6350.
20. Chen RJ, Lu MY, Wang J, Williamson DFK, Rodig SJ, Lindeman NI, Mahmood F. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Trans Med Imaging. 2022;41(4):757–70.
21. Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. Bioinformatics. 2019;35(14):446–54.
22. Vale-Silva LA, Rohr K. Long-term cancer survival prediction using multimodal deep learning. Sci Rep. 2021;11:13505.
23. Kirk PDW, Griffin JE, Savage RS, Ghahramani Z, Wild DL. Bayesian correlated clustering to integrate multiple datasets. Bioinformatics. 2012;28(24):3290–7.
24. Kim S, Kim K, Choe J, Lee I, Kang J. Improved survival analysis by learning shared genomic information from pancancer data. Bioinformatics. 2020;36(Supplement–1):389–98.
25. 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.
26. 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.
27. 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.
28. 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. 2019;16(3):841–50.
29. 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. p. 1697–700.
30. Arya N, Saha S. Multi-modal advanced deep learning architectures for breast cancer survival prediction. Knowl Based Syst. 2021;221:106965.
31. Xu J. Li W, Liu X, Zhang D, Liu J, Han J. Deep embedded complementary and interactive information for multi-view classification. In: IAAI; 2020. p. 6494–501.
32. Wang L. Chignell MH, Jiang H, Charoenkitkarn N. Cluster-boosted multi-task learning framework for survival analysis. In: IEEE international conference on bioinformatics and bioengineering; 2020. p. 255–62.
33. Erola P, Björkegren J, Michoel T. Model-based clustering of multi-tissue gene expression data. Bioinformatics. 2020;36(6):1807–13.
34. Coretto P, Serra A, Tagliaferri R. Robust clustering of noisy high-dimensional gene expression data for patients subtyping. Bioinformatics. 2018;34(23):4064–72.
35. Wang C, Guo J, Zhao N, Liu Y, Liu X, Liu G, Guo M. A cancer survival prediction method based on graph convolutional network. IEEE Trans Nanobiosci. 2020;19(1):117–26.
36. Xu X, Zhang Y, Zou L, Wang M, Li A. A gene signature for breast cancer prognosis using support vector machine. In: International conference on biomedical engineering and informatics; 2012. p. 928–31.
37. Dao F, Lv H, Wang F, Feng C, Ding H, Chen W, Lin H. Identify origin of replication in saccharomyces cerevisiae using two-step feature selection technique. Bioinformatics. 2019;35(12):2075–83.
38. Wu X, Shi Y, Wang M, Li A. CAMR: cross-aligned multimodal representation learning for cancer survival prediction. Bioinformatics. 2023;39(1):1-8.
39. Tong L, Wu H, Wang MD. Integrating multi-omics data by learning modality invariant representations for improved prediction of overall survival of cancer. Methods. 2021;189:74–85.

Publisher's Note