Medical Image Analysis 65 (2020) 101795

Contents lists available at ScienceDirect

Medical Image Analysis

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

Multi-task multi-modal learning for joint diagnosis and prognosis of human cancers

Wei Shao[a,1], Tongxin Wang[c], Liang Sun[a], Tianhan Dong[b], Zhi Han[b], Zhi Huang[d], Jie Zhang[b], Daoqiang Zhang[a,*], Kun Huang[b,*]

[a] The College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing, 211106, China
[b] The School of Medicine, Indiana University, Indianapolis, 46202, USA
[c] The Department of Computer Science, Indiana University Bloomington, Bloomington, 47405, USA
[d] School of Electrical and Computer Engineering, Purdue University, IN 47907, USA

ARTICLE INFO

Article history:
Received 14 February 2020
Revised 31 May 2020
Accepted 17 July 2020
Available online 23 July 2020

Keywords:
Multi-task multi-Modal learning
Cancer prognosis
Cancer diagnosis
Image genomics

ABSTRACT

With the tremendous development of artificial intelligence, many machine learning algorithms have been applied to the diagnosis of human cancers. Recently, rather than predicting categorical variables (*e.g.*, stages and subtypes) as in cancer diagnosis, several prognosis prediction models basing on patients' survival information have been adopted to estimate the clinical outcome of cancer patients. However, most existing studies treat the diagnosis and prognosis tasks separately. In fact, the diagnosis information (*e.g.*, TNM Stages) indicates the extent of the disease severity that is highly correlated with the patients' survival. While the diagnosis is largely made based on histopathological images, recent studies have also demonstrated that integrative analysis of histopathological images and genomic data can hold great promise for improving the diagnosis and prognosis of cancers. However, direct combination of these two types of data may bring redundant features that will negatively affect the prediction performance. Therefore, it is necessary to select informative features from the derived multi-modal data. Based on the above considerations, we propose a multi-task multi-modal feature selection method for joint diagnosis and prognosis of cancers. Specifically, we make use of the task relationship learning framework to automatically discover the relationships between the diagnosis and prognosis tasks, through which we can identify important image and genomics features for both tasks. In addition, we add a regularization term to ensure that the correlation within the multi-modal data can be captured. We evaluate our method on three cancer datasets from The Cancer Genome Atlas project, and the experimental results verify that our method can achieve better performance on both diagnosis and prognosis tasks than the related methods.

1. Introduction

Cancer is the leading cause of death in economically developed countries and the second leading cause of death in developing countries (Siegel et al., 2016). It is estimated that the number of new cases of cancer will be 539.2 per 10,000 people by the year of 2025 (Siegel et al., 2016). Thus, accurate diagnosis of cancer especially at its early stage is particularly important. So far, many biomarkers have been shown to be sensitive to the diagnosis of cancers. For example, quite a number of cancer diagnosis models (Nir et al., 2018; Coudray et al., 2018; Gecer et al., 2018) were based on histopathological images, since it can reveal the morphological characteristics of cells that are closely related to the aggressiveness of cancers. Besides histopathological images, it has been known that the genetic mutations and gene expression levels can affect the development of cancers by accelerating cell division rates (Kim and Kaelin, 2004) and modifying the tumor micro-environment (Yuan et al., 2012). Accordingly, many researchers also used genomic features such as gene expression signatures to drive diagnosis practices (Wilhelm et al., 2002; Yang et al., 2018; Niazi and Khalid, 2016). In all these diagnosis methods, classification models are learned from training samples to predict categorical variables (*e.g.*, TNM Stage) on the testing subjects.

In addition to predict categorical variables (*i.e.*, TNM Stage) as in cancer diagnosis, many prognosis prediction models were also

* Corresponding authors.
E-mail addresses: shaowei20022005@nuaa.edu.cn (W. Shao), tw11@iu.edu (T. Wang), 935800223@qq.com (L. Sun), dongtian@iu.edu (T. Dong), zhihan@iu.edu (Z. Han), zhihuan@iu.edu (Z. Huang), jizhan@iu.edu (J. Zhang), dqzhang@nuaa.edu.cn (D. Zhang), kunhuang@iu.edu (K. Huang).
[1] Wei Shao is now working in the School of Medicine, Indiana University, USA.

https://doi.org/10.1016/j.media.2020.101795
1361-8415/

adopted to perform survival analysis based on different modalities of biomarkers (Lin et al., 1993; Cheng et al., 2017a; Cooperberg et al., 2015; Li et al., 2016; Zhu et al., 2016; Yi et al., 2018; Veer and Dai, 2002). Different from the diagnosis task that focuses on the identification of current disease state, the prognosis task aims at making an prediction about the expected clinical outcome of the cancer patients. Among all the prognosis prediction models, Cox proportional hazard model (Lin et al., 1993) was the most popular one. Cheng et al. (2017a) and Cooperberg et al. (2015) used the Cox model to stratify cancer patients into subgroups with different predicted outcomes from histopathological images and genomic data, respectively. Besides the Cox model, two recent studies MTLSA (Li et al., 2016) and DeepSurv (Zhu et al., 2016), were designed to model the complex relationship between the input feature and clinical outcome. Other studies include (Yi et al., 2018) have proposed a hierarchical regression model to estimate the survival risks of different patients, and experimental results on the high-dimensional genomic data validate its superiority over the comparing methods.

Despite these progress, to the best of our knowledge, all the existing studies treated diagnosis and prognosis tasks independently, without considering the inherent correlation between them. As a matter of fact, the diagnosis information indicates the extent of the disease severity, which is highly correlated with the patients' clinical outcomes (Scarpa et al., 2010). For example, patients in stage II suffer from more aggressive cancers than those in stage I, and thus generally have higher risk for short survival time. It can be expected that better prediction performance will be achieved if we learn the diagnosis and prognosis tasks jointly, for the information of one task will be helpful to predict another related task.

At the same time, existing studies (Sun and Li, 2018; Yao and Huang, 2017; Shao et al., 2018b; Cheng et al., 2017c; Yuan et al., 2012; Huang et al., 2019) have demonstrated that the integrative analysis of images and genomic data hold great promise for cancer assessment and risk prediction. For example, Yuan et al. (2012) demonstrated that integration of lymphocyte morphology from histopathology images and gene expression signatures can significantly increase the prognosis accuracy for ER-negative breast cancer patients. Sun and Li (2018) combined the pathological images with gene expression data to classify longer-term and shorter-term survivors on breast cancer cohort. Yao and Huang (2017) developed a novel deep learning framework integrating both image and genomic data to predict the clinical outcome of cancer patients. However, direct combination of multi-modal data will increase the feature dimension, which may cause the problem of "curse of dimensionality", given the limited training samples in cancer research. Thus, feature selection, which can be considered as bio-marker identification, has become an important step for the diagnosis and prognosis of cancers. Currently, most of the existing studies (Cheng et al., 2017c; Yuan et al., 2012) first concatenated all features from histopathological images and genomic data into a long feature vector, followed by the traditional single-modality sparse learning algorithm (*e.g.*, LASSO) for key components discovery. However, these feature selection methods overlooked the correlation within the multi-modal data, which has been widely accepted as a critical component in the state-of-the-art multi-modality based machine learning methods (Liu et al., 2014; Mohammadi et al., 2017).

Inspired by the above considerations, we propose a multi-task multi-modal feature selection method (M2DP) for joint diagnosis and prognosis of cancers. Specifically, based on the task relationship learning framework, our method can automatically derive the correlation between the diagnosis and prognosis tasks, without assuming it to be known in advance. Intuitively, exploiting such task relationships can help identify a subset of bio-makers for a specific task with the knowledge of other related task. In addition, we also consider the association between different modalities by adding a regularization term to capture the inter-correlation between the selected imaging and genomic components.

To evaluate the effectiveness of the proposed method, we perform experiments on three large cancer cohorts (*i.e.*, Lung Squamous Cell Carcinoma, Breast Invasive Carcinoma, Liver Hepatocellular Carcinoma) in The Cancer Genome Atlas (TCGA). The experimental results verify that our proposed M2DP method not only can achieve better performance on both diagnosis and prognosis tasks than competing algorithms, but also help to discover useful histopathological image and genomic bio-markers for the prediction of the development of cancers.

Our preliminary work that only using the diagnosis information can help achieve better prognosis prediction performance was published in MICCAI 2019 (Shao et al., 2019). In this substantially expanded journal paper, we offered new contributions in the following aspects: 1) further demonstrating that the proposed model can also promote the prediction performance for diagnosis tasks; 2) evaluating the effectiveness of the proposed method on two additional datasets (*i.e.*, Breast Invasive Carcinoma and Liver Hepatocellular Carcinoma datasets); 3) in-depth explanation of the biomarkers that are identified by the proposed model. 4) visualization of the selected image features in both high and low survival risk groups. 5) discussion on effects of parameter σ in the proposed M2DP model.

2. Materials and methods

2.1. Datasets.

The Cancer Genome Atlas (TCGA) is a large consortium project that has generated genomic and imaging data for thousands of tumor samples across more than 30 types of cancers (Zhu et al., 2014). In this study, we test our method on three early-stage (*i.e.*, stage I and stage II) cancer cohorts including Lung Squamous Cell Carcinoma (LUSC), Breast Invasive Carcinoma (BRCA) and Liver Hepatocellular Carcinoma (LIHC) from TCGA, since the diagnosis and prognosis of early-stage cancer patients are critical for cancer treatment but still challenging (Cheng et al., 2017c). For each cancer cohort, we only collect patients whose histopathological image, mRNA, diagnosis information (*i.e.*, TNM stage) and clinic outcome (*i.e.*, survival status and time) are all available. The details of the cohort information are listed in Table 1, where censored patients mean that these patients did not suffer the outcome of death events during the follow-up period, while the non-censored category refers to the patients, whose survival information are accurate from diagnosis to death. In addition, in order to find out the association between censoring and stage information in different cancer cohorts, we also calculated the number of censored/non-censored patients that belong to stage I and stage II categories in Table 2.

As can be seen from Table 2, the Stage I patients are more likely to be censored patients. Specifically, for LIHC and BRCA datasets, the ratios of censored patients in Stage I ($112/137 = 0.818$, $164/180 = 0.911$) are higher than those in Stage II ($40/55 = 0.727$, $544/610 = 0.891$). While for LUSC cohort, the ratio of censored patients in Stage I ($121/205 = 0.590$) is only slightly smaller than that in stage II category ($82/137 = 0.599$). One possible reason is that the patients with stage I are generally in the early-stage of

Table 1
Patient demographics for different cancer datasets.

	LUSC	BRCA	LIHC
Censored / Non-Censored	203/139	708/82	152/40
Stage I / Stage II	205/137	180/610	137/55
Total	342	790	192

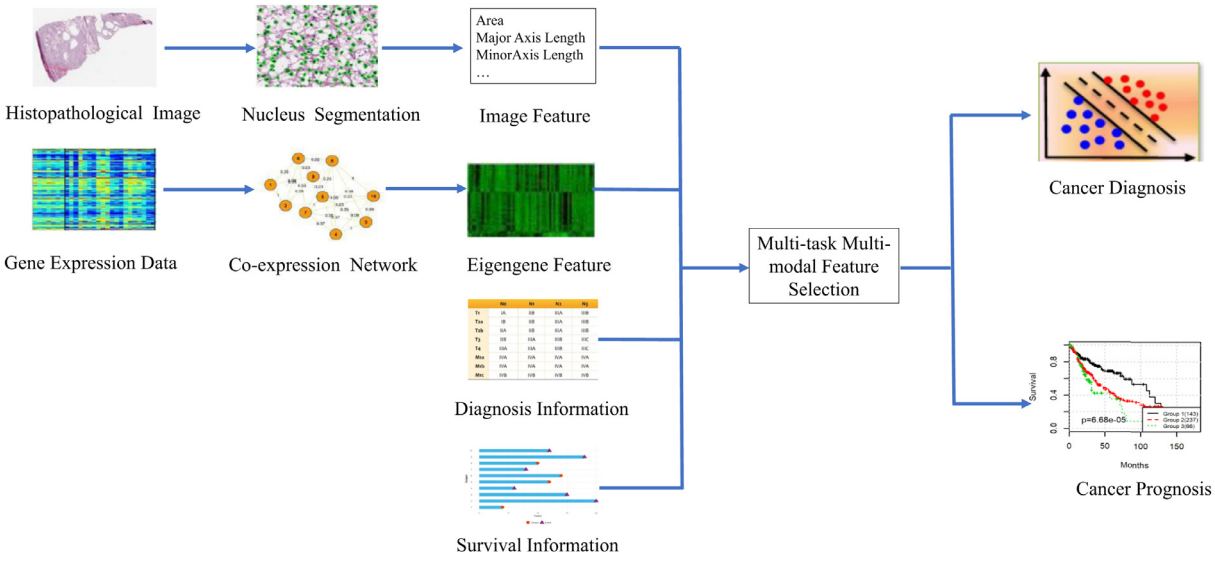

Fig. 1. The flowchart of the proposed method.

Table 2
The number of censored/non-censored patients that belong to stage I and stage II categories for different cancer cohorts.

	LIHC		BRCA		LUSC	
	Stage I	Stage II	Stage I	Stage II	Stage I	Stage II
Censored	112	40	164	544	121	82
Non-Censored	25	15	16	66	84	55

the cancer disease, and their death event may not be observed for the given period.

2.2. Overview of our method.

We summarize our framework in Fig. 1, which consists of the following three steps. Firstly, extracting imaging and eigengene features from the histopathological image and gene expression data, respectively. Secondly, implementing the proposed multi-task multi-modal feature selection algorithm (*i.e.,* M2DP) to identify diagnosis and prognosis related features. Thirdly, based on the selected features of each patient, applying AdaBoosting and Cox proportional hazard models for the diagnosis and prognosis prediction of cancer patients, respectively.

2.3. Feature extraction.

For the diagnosis and prognosis of cancers, different types of bio-markers (*i.e.,* image and genomic data) are expected to provide complementary information. Therefore, features from both histopathological images and gene expression data are extracted in our proposed method. Specifically, for whole-slide images of each patient, 2–8 regions of interest (ROI) of size 3K by 3K are extracted by an expert at first. Then, we apply an unsupervised method introduced in Phoulady and Dmitry (2016) to segment the nuclei from these ROIs. Next, for each segmented nucleus, we extract seven cell-level features characterizing the nuclei area (denoted as area), the major and minor axis length of cell nucleus(major and minor), the ratio of major axis length to minor axis (ratio), and mean, maximum, and minimum distances to its neighbouring nuclei (distMean, distMax, and distMin). Finally, for each cell-level feature, we summarized all the cell-level features from the images of the same patient (a patient may have more than one slides) using a 10-bin histogram and five statistical measurements (*i.e.,* mean, standard deviation, skewness, kurtosis, and entropy). Thus a 105-dimensional imaging feature can be derived for each patient. We use the same naming rule for both cell-level and patient-level features. For instance, the features 'major_1' represents the percentage of cells with small major axis length while 'major_10' refers to the percentage of nuclei with long major axis.

As to genomic features, instead of directly examining data with the large number of genes, which contain noise and poses a challenge to obtain sufficient statistical power (Cheng et al., 2017c), we reduce the data dimensionality by mining gene co-expression networks (GCNs). Specifically, we identify densely connected GCN Modules by applying a weighted network mining algorithm called local maximum quasi-clique merging (lmQCM)Zhang and Huang (2016) and summarize each Module as an eigengene. The lmQCM has previously been demonstrated to be able to identify smaller yet tightly co-expressed gene modules comparing to the commonly used WGCNA algorithm (Zhang and Horvath, 2005). The eigengene value of a Module is defined as its first principle component (Zhang and Huang, 2016), which can be considered as a representative gene expression profile in that Module. This process yields 58, 50 and 60 eigengenes for LUSC, BRCA and LIHC datasets respectively.

2.4. Multi-task relationship learning.

In this paper, we focus on the diagnosis and prognosis prediction of cancer patients from multi-modal data. Since the diagnosis information indicates the extent of the disease severity that is highly correlated with the patients' clinical outcomes, we aim to seek a jointly multi-task learning framework (MTL) to automatically discover the relationship between the diagnosis and prognosis tasks that can boost the prediction performance for both tasks. In the following, we will elaborate the multi-task relationship learning model in detail.

Let $X = [x_1, x_2, \ldots, x_N]^T = [X^H, X^G] \in R^{N\times(p+q)}$. Here, $X^H \in R^{N\times p}$ and $X^G \in R^{N\times q}$ correspond to the histopathological imaging data and eigengene data, respectively. N is the number of the patients, and p and q are the feature number of imaging data and eigengene data, respectively. For the diagnosis task, we use (x_i, y_i) to represent each observation, where $x_i \in R^{p+q}$ is the patient feature vector and $y_i \in \{1, 2\}$ refers to the categorical TNM information(*i.e.*, stage I and stage II). As to the prognosis task, each sample is denoted by

the triplet (x_i, t_i, δ_i), where t_i is the observed survival time, and δ_i is the censoring indicator. Here, $\delta_i = 1$ or $\delta_i = 0$ indicates a non-censored or a censored instance, respectively. Then, the objective function of the MTL framework is:

$$\begin{aligned} &\min \sum_{k\in\{S,D\}} \sum_{i=1}^{N} l_k(x_i, w_k) + \gamma \,\mathrm{tr}(\mathrm{W}\mathrm{M}^{-1}\mathrm{W}^{\mathrm{T}}) \\ &s.t. \quad M \geq 0, tr(M) = 1 \end{aligned} \tag{1}$$

In the first term of Eq. (1), $w_D \in R^{p+q}$ and $w_S \in R^{p+q}$ correspond to the linear discriminant functions for the diagnosis and prognosis tasks on the multi-modal data, respectively. $l_D(x_i) = \|x_i w_D - y_i\|_2^2$ indicates the empirical loss for the diagnosis task, while $l_S(x_i) = c_i \|x_i w_S - t_i\|_2^2$ is a weighted regression term for prognosis task, where c_i is the weight of the survival regression loss for each patient defined as follows:

$$c_i = \begin{cases} 1 & if\ \delta_i = 1 \\ 0 & if\ \delta_i = 0\ and\ x_i w_S - t_i > 0 \\ \sigma & if\ \delta_i = 0\ and\ x_i w_S - t_i \leq 0 \end{cases} \tag{2}$$

where $c_i = 1$ if x_i is a non-censored patient. By considering that the actual survival time for a censored instance should be larger than its observed time, we define c_i as 0 if $x_i w_S - t_i \geq 0$. Otherwise, we let c_i equal to a constant $\sigma = 1.5$ that is larger than 1 since the difference between actual survival time and the estimated survival time is indeed greater than $t_i - x_i w_S$.

The second term of Eq. (1) is used to capture the inherent correlation between different tasks, where $W = [w_D, w_S] \in R^{(p+q)\times 2}$ and M^{-1} denotes the inverse of the matrix $M \in R^{2\times 2}$. Here, M is defined as a task covariance matrix (Zhang and Yeung, 2010) that will benefit learning on W by automatically inducing the correct relationship M between diagnosis and prognosis tasks. The constraint term $M \geq 0$ is used to restrict M as a positive semidefinite matrix, and $tr(M) = 1$ is used to penalize the complexity of M.

2.5. Multi-task multi-modal feature selection for joint diagnosis and prognosis of cancers.

In the multi-task learning (MTL) framework, simply concatenating the derived imaging and genomic feature together may bring the problem of "curse of dimensionalities". Thus, feature selection, which can be considered as the identification of key bio-markers, should be applied to select useful features. In addition, the above MTL model only considers the association between diagnosis and prognosis tasks, but overlooks the correlation between the imaging and genomic data. As a matter of fact, the exploitation of multi-modal association has been widely accepted as a key component of current multi-modality based machine learning approaches (Liu et al., 2014; Mohammadi et al., 2017). Based on the above considerations, we propose a novel multi-task multi-modal feature selection algorithm (M2DP) for joint diagnosis and prognosis of cancers. Specifically, we divide w_D and w_S in Eq. (1) into $w_D = [w_D^H, w_D^G]^T$ and $w_S = [w_S^H, w_S^G]^T$, where $w_D^H \in R^p$ and $w_S^H \in R^p$ correspond to the linear function for imaging feature, while $w_D^G \in R^q$ and $w_S^G \in R^q$ refer to the linear function for genomic feature. Then, the objective function of the M2DP model is formulated as:

$$\begin{aligned} &\min_{W,M} \sum_{k\in\{S,D\}} \sum_{i=1}^{N} l_k(x_i) + \alpha \sum_{k\in\{S,D\}} \left\| X^H w_k^H - X^G w_k^G \right\|_2^2 \\ &\quad + \sum_{k\in\{S,D\}} \sum_{j\in\{H,G\}} \beta_k^j \left\| w_k^j \right\|_1 + \gamma tr(WM^{-1}W^T) \\ &\quad s.t. \quad M \geq 0, tr(M) = 1 \end{aligned} \tag{3}$$

Where the first and fourth terms in Eq. (3) are the same as they are stated in the MTL framework (shown in Eq. (1)). The second term is used to enforce the projections of each modality (*i.e.*, $X^H w_k^H$ and $X^G w_k^G$) to get as close as possible. So that the inter-correlations across different modalities can be captured. The third L1-norm regularized terms are used to select a small number of features from the multi-modal data for different tasks. Here, β_k^j $(k \in \{S, D\}, j \in \{H, G\})$ are regularization parameters and larger β_k^j value implies that fewer features in the j-th modality will be preserved for the prediction task k.

2.6. Optimization.

We adopt an alternating strategy to optimize W and M in the proposed M2DP model. Specifically, given a fixed M, we define $M^{-1} = [M_{DD}, M_{DS}; M_{DS}, M_{SS}] \in R^{2\times 2}$. Then, the optimization problem for W is:

$$\begin{aligned} \min_W &\left\| C(X^H w_S^H + X^G w_S^G - T) \right\|_2^2 + \left\| X^H w_D^H + X^G w_D^G - Y \right\|_2^2 \\ &+ \sum_{k\in\{S,D\}} \sum_{j\in\{H,G\}} (\beta_k^j \left\| w_k^j \right\|_1 + \gamma M_{kk} (w_k^j)^T w_k^j) \\ &+ 2rM_{DS} \sum_{j\in\{H,G\}} (w_D^j)^T w_S^j + \alpha \sum_{k\in\{S,D\}} \left\| X^H w_k^H - X^G w_k^G \right\|_2^2 \end{aligned} \tag{4}$$

Where $T = [t_1, t_2, \ldots, t_N]^T \in R^N$ and $Y = [y_1, y_2, \ldots, y_N]^T \in R^N$ indicate the recorded survival time and TNM stage information for the patients in the training set. $C \in R^{N\times N}$ is a diagonal matrix, with the i-th element defined as c_i in Eq. (2). It is worth noting that, Eq. (4) is convex with respect to each w_k^j $(k \in \{S, D\}, j \in \{H, G\})$ in W, and thus can be solved alternatively to obtain the optimal W. Here, we only provide the optimization procedure for w_D^H, other variables could be derived using similar way. Specifically, when w_D^G, w_S^G, w_S^H is given and fixed, the objective function for finding w_D^H could be stated as:

$$\begin{aligned} \min_{w_D^H} &\left\| X^H w_D^H + X^G w_D^G - Y \right\|_2^2 + \beta_D^H \left\| w_D^H \right\|_1 + \gamma \mathrm{M}_{DD} (w_D^H)^T w_D^H \\ &+ 2rM_{DS} (w_D^H)^T w_S^H + \alpha \left\| X^H w_D^H - X^G w_D^G \right\|_2^2 \end{aligned} \tag{5}$$

Since the L1-norm is non-differentiable at zero, a smooth approximation is estimated for L1 term by including an extremely small value. Then, by taking the derivative with respect to w_D^H and let it to be zero, we can obtain:

$$\begin{aligned} w_D^H = &[(2 + 2\alpha)(X^H)^T X^H + \beta_D^H K_D^H + 2rM_{DD} I]^{-1} \\ &\times [(2\alpha - 2)(X^H)^T X^G w_D^G + 2(X^H)^T Y - 2rM_{DS} w_S^H] \end{aligned} \tag{6}$$

where K_D^H is a diagonal matrix with the d-th element denoted as $\frac{1}{\|(w_D^H)_d\|_1 + \xi}$. Here, $(w_D^H)_d$ represents the d-th elements in w_D^H, and ξ is a very small positive real number (*i.e.*, $\xi = 1e-3$).

After each $w_k^j (k \in \{S, D\}, j \in \{H, G\})$ in W is determined, we follow the method in Zhang and Yeung (2010) and Liu et al. (2015) to get the closed-formed solution for M as follows:

$$M = (W^T W)^{\frac{1}{2}} / tr((W^T W)^{\frac{1}{2}}) \tag{7}$$

Finally, we renew the weight of each instance *i.e.*, c_i according to Eq. (2). The above steps are repeated until W, M converge. We show the optimization algorithm for the proposed M2DP method in Table 3.

2.7. Diagnosis and prognosis of human cancers.

For the selected multi-modal feature of diagnosis tasks, we simply apply the AdaBoosting algorithm to classify patients into stage I or stage II category. Here, we fix the number of weak learner as 200 to train the classifier and the learning rate is set as 0.001. As to prognosis task, the Cox proportional hazard model (Lin and Wei, 1989) is applied to predict the patients' clinical outcome, and

Table 3
The optimization algorithm for the proposed M2DP method.

The optimization procedure for M2DP method.
Input:
$X^i (i \in \{H, G\})$ % *The data matrix for multi-modal data.*
$Y \in R^N$ % *The diagnosis information.*
$T \in R^N$ % *The recorded survival time.*
Initialize:
$w_k^j = 0$ $(k \in \{S, D\}, j \in \{H, G\})$ *%Initialized as zero vector.*
$M = [1, 0; 0, 1]$ *%Initialized as identity matrix.*
While not converge do:
Update w_k^j $(k \in \{S, D\}, j \in \{H, G\})$ in turn according to Eq. (6);
Update M according to Eq. (7);
Update C according to Eq. (2);
End While
Output:
$w_k^j (k \in \{S, D\}, j \in \{H, G\})$ % *The non-zero elements in w_k^j correspond to the selected multi-modal feature for different tasks.*

its negative log partial likelihood function can be formulated as Lin and Wei (1989):

$$l(\theta) = \sum_{i=1}^{N} \delta_i \left(\theta^T x_i - \log \sum_{j \in R(t_i)} \exp(\theta^T x_j) \right) \tag{8}$$

Where $R(t_i)$ is the risk set at time t_i, which represents the set of patients that are still under risk before time t. It is worth noting that $f(x) = \theta^T x$ is also being called as risk function, and we can estimate the survival risks of different patients after θ is determined. To evaluate the prognosis performance of different methods, we adopt Concordance Index (CI)Chen et al. (2012) and Brier Score(BS) Wang et al. (2019) as evaluation metrics. Here, the formulation of the CI is shown as follows:

$$CI = \frac{1}{n} \sum_{i \in \{1 \ldots N | \delta_i = 1\}} \sum_{t_j > t_i} I(x_i \theta > x_j \theta) \tag{9}$$

Where n denotes the number of comparable pairs, t_i corresponds to the actual survival observation for the *i-th* patient, $I(.)$ is the indicator function, and $\delta_i = 1$ indicates an non-censored instance. It is worth noting that the CI index measures the fraction of all pairs of patients whose survival risks are correctly ordered. The CI value ranges from 0 to 1, where larger CI values indicate the better prognosis prediction performance. On the other hand, the measurement of Brier Score(BS) represents the average squared distances between the observed survival status and the predicted survival probability and is always a number between 0 and 1, with 0 being the best possible value.

3. Experimental results

3.1. Experimental settings.

To evaluate the performance of the proposed M2DP method for the diagnosis and prognosis of cancer patients, we test it on three early-stage cancer cohorts (*i.e.*, LUSC, BRCA and LIHC) derived from the TCGA database. For each cohort, we randomly partition it into 5 folds. Here, we enforce the ratio of Stage I and Censored patients in each fold approximate to that in the whole cohort with at most $\pm$ 0.05 gap, and we show the ratio of Stage I and censored patients in each fold in Tables S4-S6 in the Supplementary Materials. At each time, subjects within one fold are used as testing data, while all the other subjects in the remaining 4 folds are used for model training. For the parameter settings of the M2DP model (shown in Eq. (3)), $\beta_k^j (k \in \{S, D\}, j \in \{H, G\})$ are tuned from $\{2^3, 2^4, 2^5, 2^6\}$, α and γ are tuned from $\{2^4, 2^5\}$ and $\{2^3, 2^4, 2^5\}$, respectively. For different modality of data, a common feature normalization strategy is adopted. Specifically, for each feature f_i in the training set, the normalized feature is $f_i' = (f_i - m_i)/\epsilon_i$, where m_i and ϵ_i correspond to the mean and standard deviation of f_i. Also, the m_i and ϵ_i values are used to normalize its corresponding feature in the testing set.

3.2. Learned relationship between diagnosis and prognosis tasks.

In this section, we visualize the learned correlation coefficient matrix (*i.e., M* in Eq. (3)) between the diagnosis and prognosis tasks for different cancer cohorts. Since the correlation coefficients are different for each fold of the cross-validation, we have visualized the mean and variance of the correlation coefficients for each cohort in Fig. 2, where green and yellow in the color bar denote a high positive correlation coefficient. From Fig. 2, we observe that the diagnosis and prognosis tasks are all positively correlated in each cancer cohort, which is consistent with the fact that the disease severity is associated with the patients' clinical outcome. In addition, we also find that the variance of the correlation coefficients for LIHC, BRCA and LUSC datasets are 0.057,0.034,0.042 respectively, which demonstrate that the correlation between diagnosis and prognosis tasks is consistent in each test fold. Hence, the exploration of the real relationships among multiple tasks via the regularization term in Eq. (3) can help to improve the learning performance of both diagnosis and prognosis tasks.

3.3. Comparison of M2DP with other methods for the diagnosis of cancer patients.

In this section, we compare the diagnosis power of the proposed M2DP with the following 4 algorithms on different datasets

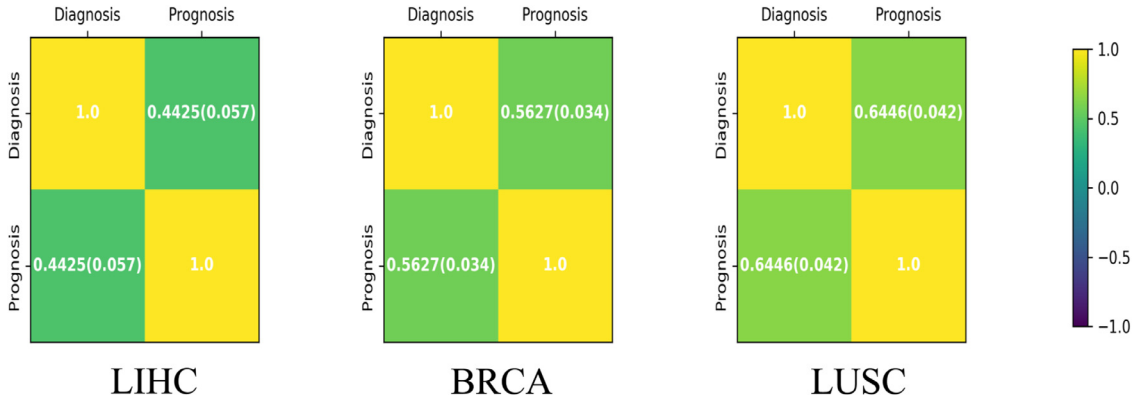

Fig. 2. The mean and variance of the correlation coefficients between the diagnosis and prognosis tasks for LIHC,BRCA and LUSC datasets by M2DP algorithm.

Table 4
The average classification performance of different methods in diagnosing the cancer patients on different cancer cohorts.

Dataset	Method	S-LASSO	M-LASSO	DSCCA	MTFS	M2DP (ours)
LIHC	Accuracy (%)	62.73 ± 5.31	64.38 ± 3.07	66.41 ± 6.61	71.29 ± 2.99	**74.25 ± 3.84**
	Recall (%)	72.67 ± 4.27	80.86 ± 3.15	79.41 ± 6.87	84.53 ± 2.19	**87.51 ± 2.51**
	Precision (%)	64.82 ± 5.35	68.92 ± 3.87	70.26 ± 3.76	73.58 ± 4.75	**75.45 ± 4.15**
	F1-Score (%)	69.32 ± 4.84	74.41 ± 3.66	74.56 ± 4.38	78.68 ± 3.66	**81.20 ± 3.12**
BRCA	Accuracy (%)	59.44 ± 3.21	65.32 ± 2.89	64.43 ± 2.15	70.47 ± 5.62	**72.53 ± 3.65**
	Recall (%)	68.48 ± 5.92	77.45 ± 5.45	79.06 ± 5.84	82.72 ± 3.20	**88.18 ± 1.03**
	Precision (%)	64.00 ± 4.70	68.22 ± 2.69	72.98 ± 3.74	77.67 ± 4.33	**78.76 ± 4.10**
	F1-Score (%)	66.01 ± 5.24	74.11 ± 3.01	75.43 ± 4.21	79.34 ± 3.77	**83.17 ± 2.55**
LUSC	Accuracy (%)	56.79 ± 5.05	65.83 ± 3.69	66.73 ± 4.15	67.57 ± 4.97	**70.08 ± 3.01**
	Recall (%)	66.28 ± 3.92	74.34 ± 6.75	76.17 ± 5.84	79.23 ± 3.21	**81.25 ± 3.73**
	Precision (%)	60.35 ± 4.67	67.46 ± 5.69	69.27 ± 4.74	**74.29 ± 5.69**	73.57 ± 4.33
	F1-Score (%)	64.22 ± 4.02	71.25 ± 5.72	73.01 ± 5.04	76.54 ± 3.84	**77.49 ± 4.23**

by the measurements of *Accuracy*, *Recall*, *Precision* and *F1-Score*: 1) S-LASSO: Using LASSO with single modality data (*i.e.*, imaging or genomic data), where the modality yielding to the best accuracy is reported. 2) M-LASSO: Concatenate the multi-modal data into a long vector, and use the LASSO algorithm for feature selection. 3) DSCCAWitten and Tibshirani (2009): a supervised multi-modal feature selection method for cancer diagnosis without the guidance of survival information. 4) MTFS: A variant of M2DP, which overlooks the correaltion between different modalities (*i.e.*, the second term in Eq. (3)). For the fair of comparison, all the comparing feature selection algorithms are followed by the AdaBoosting algorithm for cancer diagnosis. The results are shown in Table 4.

From Table 4, we can derive the following observations: 1) The algorithms (*i.e.*, M-LASSO, DSCCA and MTFS) combining both imaging and genomic data can consistently achieve more accurate discrimination than the single-modality based algorithm (*i.e.*, S-LASSO) on all of the three datasets. These results clearly verify the advantage of the integrative analysis of pathological image and genomic data for cancer diagnosis. 2) The proposed M2DP and its variant MTFS can achieve better classification performance than their competitors. These results validate the superiorities of capturing the correlation between diagnosis and prognosis tasks can help promote the diagnosis performance. 3) The proposed M2DP can better diagnose cancer patients with different TNM stages than the MTFS method, which also shows the advantage of taking the correlation among different modalities into account for feature selection.

3.4. Comparison of M2DP with other methods for the prognosis prediction of cancer patients.

For the prognosis prediction of cancer patients, we compare M2DP with the following six algorithms using Concordance Index (CI) and Brier Score(BS) that are introduced Section 2.7. 1) Stage: Only use the stage information to select useful features for survival analysis. 2) SLASSO-COX(SCOX): The best prognosis prediction result on individual modality of data by applying the LASSO-COX modelRobert (1997). 3) MLASSO-COX(MCOX): Simply concatenate the image and eigengene feature together, and then use LASSO-COX model for prognosis prediction. 4) DMLASSO-COX(DMCOX): Add the diagnosis information as a feature to the multi-modal data, then use the LASSO-Cox model for patient outcome prediction. 5) OSCCAShao et al. (2018a): A multi-modal feature selection method for survival analysis without the guidance of diagnosis knowledge. 6) MTFS: A variant of M2DP, which miss the second term in Eq. (3) that can capture the correlation between imaging and eigengene feature. The experimental results are shown in Tables. 5 and 6.

As can be seen from Tables 5 and 6, M2DP consistently outperforms the comparing methods on all of the three cancer cohorts. Specifically, our proposed method M2DP achieves the best average Concordance Index (CI) of 0.696, 0.726 and 0.727 for LIHC, BRCA and LUSC, respectively, and the Brier Score (BS) of our method is also superior to its competitors on all of the three cancer cohorts. These results clearly demonstrate that the incorporation of the prior diagnosis information (*i.e.*, TNM stage) and considering

Table 5
Comparison of different feature selection methods on Concordance Index(CI) for prognosis prediction, the larger CI indicates better prognosis results.

Dataset	Stage	SCOX	MCOX	DMCOX	OSCCA	MTFS	M2DP
LIHC	0.547 (±0.07)	0.595 (±0.08)	0.569 (±0.05)	0.654 (±0.04)	0.643 (±0.06)	0.663 (±0.06)	0.696 (±0.05)
BRCA	0.551 (±0.09)	0.588 (±0.09)	0.648 (±0.07)	0.689 (±0.04)	0.672 (±0.06)	0.691 (±0.07)	0.726 (±0.06)
LUSC	0.584 (±0.06)	0.597 (±0.04)	0.647 (±0.05)	0.667 (±0.04)	0.687 (±0.04)	0.707 (±0.06)	0.727 (±0.07)

Table 6
Comparison of different feature selection methods on Brier Score(BS) for prognosis prediction, the smaller BS indicates better prognosis results.

Dataset	Stage	SCOX	MCOX	DMCOX	OSCCA	MTFS	M2DP
LIHC	0.342 (±0.06)	0.317 (±0.05)	0.294 (±0.06)	0.282 (±0.04)	0.234 (±0.04)	0.243 (±0.05)	0.199 (±0.02)
BRCA	0.416 (±0.04)	0.372 (±0.06)	0.295 (±0.05)	0.286 (±0.02)	0.235 (±0.04)	0.184 (±0.06)	0.171 (±0.02)
LUSC	0.398 (±0.06)	0.366 (±0.05)	0.327 (±0.04)	0.289 (±0.07)	0.275 (±0.04)	0.252 (±0.03)	0.231 (±0.02)

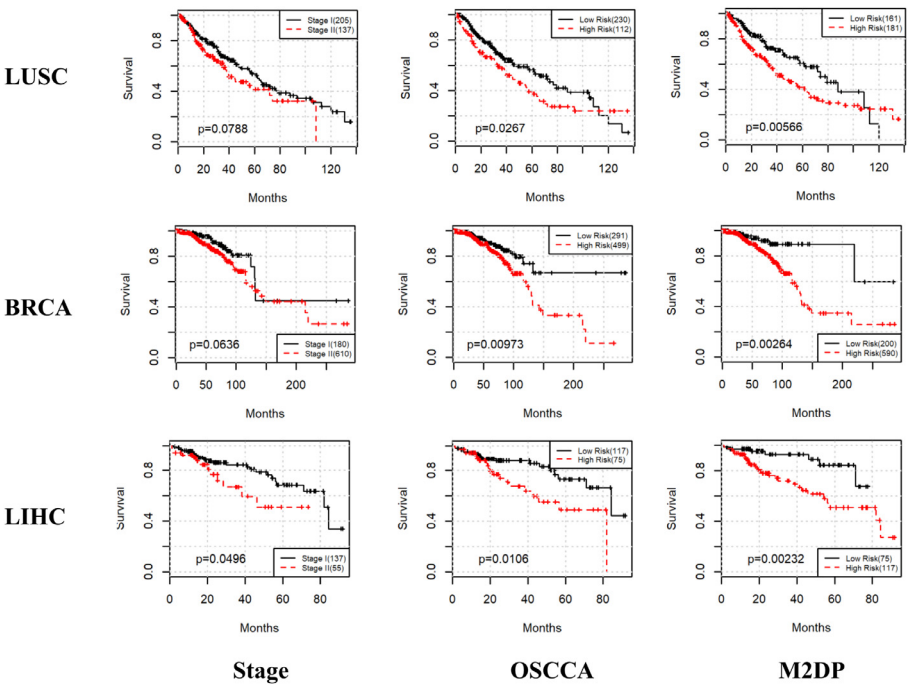

Fig. 3. The survival curves by applying different methods on LUSC, BRCA and LIHC datasets.

the association within the multi-modal data can help improve the prognosis performance. In addition, both the experimental results shown in Table 4, Table 5 and Table 6 indicate that the diagnosis and prognosis tasks are highly correlated with each other, and applying the multi-task learning framework to capture the correlation between different tasks can promote the prediction performance for each individual task.

3.5. Comparison of M2DP with other methods for patient stratification.

In the field of prognosis prediction, another important task is to stratify cancer patients into subgroups with different predicted outcomes, by which we could intervene with personalized treatment plan in the stage of cancer development. In this section, we also compare the prognostic power of different approaches by stratifying cancer patients into 2 subgroups (*i.e.*, the high and low survival risk groups) with different predicted outcomes, with the experimental results shown in Fig. 3. Specifically, the stage method divides all the patients into 2 subgroups according to the TNM stage (*i.e.*, stage I and stage II). For OSCCA (Shao et al., 2018a) and the proposed M2DP approaches, k-means clustering algorithm is adopted to aggregate the patients into 2 subgroups based on the selected features. Then, we test if these 2 subgroups has significantly different survival outcome using log-rank test (Wei et al., 1990).

As can be seen from Fig. 3, the proposed M2DP method can achieve superior stratification performance (*i.e.*, with the p-values of $5.66e-3, 2.64e-3$ and $2.32e-3$ for LUSC, BRCA and LIHC dataset, respectively) than its competitors. These results further show the promise of applying task relationship learning framework to select survival-associated feature for patient stratification from multi-modal data, which opens up the opportunity for building personalized treatment plan in the stage of cancer development.

3.6. Significant biomarkers detected by the proposed M2PD method.

It is of great importance to identify potential bio-markers that are associated with the diagnosis and prognosis of cancers. In this study, we evaluate the discriminative power of the selected multi-modal features (*i.e.*, image and eigengene features) using the proposed M2PD method. As the selected features in each task (*i.e.*, diagnosis and prognosis tasks) and each fold of cross-validation are different, we list common features and their corresponding weights that are appeared in at least four folds of the five-fold cross-validationin in both tasks as the most discriminative features. The experimental results are shown in Table 7.

Table 7
The selected features and their corresponding weights by applying M2DP model.

DataSet	Feature Type	Selected Features
LIHC	Eigengene	Module 14(0.141), Module 18(0.042), Module 19(0.027), Module 22(0.043), Module 28(0.021), Module 43(0.031), Module 46(0.024)
	Image	Major_10(0.321), Major_kurtosis (0.034), Minor_1 (0.021), Minor_std (0.121), Ratio_10 (0.127), Ratio_2 (0.071), Ratio_8 (0.05), Ratio_9 (0.041), distMin_9 (0.021)
BRCA	Eigengene	Module 2(0.551), Module 4(0.362), Module 10(0.081), Module 14(0.043),Module 20(0.110), Module 25(0.021) Module 37 (0.017), Module 40(0.042), Module 46 (0.085)
	Image	Area_1(0.012), Ratio_2(0.110), distMin_5(0.081), Ratio_1(0.051), distMin_3(0.132), distMin_skewness(0.017), Major_3(0.041), Major_9(0.096), distMin_2(0.092), distMean_4(0,071), Major_kurtosis (0.043).
LUSC	Eigengene	Module 7(0.022), Module 10(0.026), Module 19(0.013), Module 20(0.008), Module 29(0.038), Module 37(0.028), Module 45(0.017).
	Image	distMean_2(0.026), Major_1(0.038), Major_4 (0.034), Major_1(0.016), Minor_3(0.041), distMin_kurtosis(0.014), Minor_5(0.014)

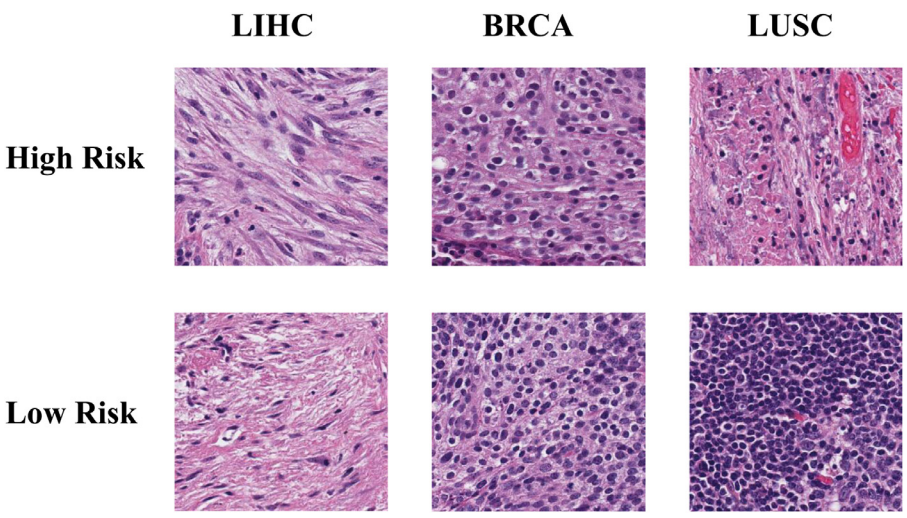

Fig. 4. Sample patches that are associated with the selected image features in both high and low survival risk groups.

Moreover, as can be seen from Table 7, nine image features are identified for the LIHC cohort. Among these identified image features, Major_10, Ratio_10, Ratio_8, Ratio_9 and Minor_1 correspond to the spindle-shaped stromal cells (*i.e.*, fibroblasts) that are characterized by relatively large ratio between major and minor axes, and it is known that the percentage of stromal tissues plays an important role in the development of Liver cancer Affo et al. (2017). For Eigengene features, our model selects 7 genomic co-expression Modules (see Table S1 in the Supplementary Materials for details). The enrichment analysis by the Toppgene (https://www.toppgene.cchmc.org/) shows that Module 14 and Module 18 are enriched with genes involved in the development extracellular matrix with the p-values of $2.38e-4$ and $1.36e-3$, respectively. Both of the selected image and eigengene features indicate that, besides the unlimited growth of tumor cells, tumor stroma in the microenvironment also plays an important role in the development of Liver cancers Cheng et al. (2017b). In addition, Module 43 is consisted of 12 genes, where CLCA4 in that module can inhibit cell proliferation and invasion of liver cancer by suppressing epithelial-mesenchymal transition via PI3K/AKT signaling (Liu et al., 2018) and GPA33 is an important biomarker that usually overexpresses in primary hepatocellular carcinoma (Ai et al., 2017).

For BRCA cohort, eleven image features are selected. The cells with small size, round shape and small neighboring distance (*i.e.*, Major_3, Ratio_2, Ratio_1, Area_1 and distMin_3) usually correspond to the lymphocytes that cluster together in the tissue images. As to the genomic feature of BRCA, Module 2 (details are shown in Supplementary Table S2) contains 208 genes and 107 of them are enriched in the regulation immune system process (with the p-value of $4.77e-62$), and 12 genes in Module 10 are also associated with human immnue response (with the p-value of p-value $= 1.368e-8$). In addition, Module 40 is mostly comprised of HLA genes (*i.e.*, HLA-F,HLA-B,HLA-A,HLA-H and HLA-E) that are highly enriched with cytotoxic T cell activation in the immune process. It is worth noting that the identified image features and gene Module 2, Module 10 and Module 40 are all strongly associated with immune response, which plays an important role in the development and treatment of breast cancer (Denkert et al., 2010). Moreover, the enrichment analysis on Module 20 shows that 9 of 16 genes (*i.e.*, ARFGEF1, CPNE3, FAM91A1, MTDH, RADN1, AZIN1, UTP23, ZNF252P, ZNF623) in the module are involved in female cancers (See Supplementary Table S2 for details), and Module 37 is consisted of 11 genes, and three of them (i.e., UQCRHL, NDUFS5, UQCRH) are involved in the pathway of oxidative phosphorylation that are downregulated in human cancers (Ashton et al., 2018).

As to the LUSC cohort, the identified image features (*i.e.*, Minor_3, distMean_2) also refer to the small-size and high-density lymphocytes, which is in accordance with the fact that the activity of lymphocytes is associated with the diagnosis and prognosis of lung cancer patients (Liu and Zeng, 1997). In addition to image features, our M2DP model selects seven Eigengene Modules (details are shown in Table S3 in the Supplementary Materials.). Specifically, two genes (i.e., RBM5 and RBM 6) in Module 7 have been demonstrated to be associated with the cancer cell proliferation in lung cancer (Lerman and Minna, 2000). Module 10 involves 30 genes, and nine of them are enriched with adaptive immune response to cancers (p-value=$4.51e-7$). Furthermore, two genes (TNFRSF17, CD27) in Module 37 have been shown to specifically bind to the tumor necrosis factor that play a central role in immune homeostasis (Cho et al., 2004), while 8 of 13 genes in Module 29 are enriched with the biological process of T-cell activation (p-value=$1.09e-10$) in the immune response, and it has been reported that immune response plays a pivotal role in the development and treatment of lung cancers and the progression of Lung cancer is associated with the increased dysfunction of T-cells (Thommen et al., 2015).

Last but not least, we also visualize the sample patches that are associated with the selected image feature in both high and low survival risk groups (*i.e.*, shown in Fig. 4). As can be seen from Fig. 4, the first column shows the visualization of the image feature Major_10 and Ratio_10 for LIHC dataset. Here, the large ratio between major and minor axes correlates to the stromal cells, and the patients in high-survival risk group have larger percentage of stromal cells than those in the low-survival risk group. Those observations are consistent with the knowledge that a high proportion of stromal cells can assist the cancer cells to colonize the micro-environment and form metastasis, and thus is associated with poor clinical outcome (Wu and Su, 2016). The second and third columns visualize the features (*i.e.*, Major_3, distMin_3 and Ratio_2 for BRCA cohort, Minor_3 and distMin_2 for LUSC cohort) that are related to lymphocytes in the cancer tissues. From the second and third columns of Fig. 4, we observe that the high percentage of lymphocytes will result in the low survival risk, which is concordance with the fact that the high lymphocyte infiltration in cancer tissues indicates stronger immune responses to cancer cells, leading to better clinical outcome (Cheng et al. 2017c).

Table 8
The Effect of the Parameter σ for the Diagnosis and Prognosis of Cancers.

Method	$\sigma = 0.5$		$\sigma = 1.5$		$\sigma = 2$		$\sigma = 2.5$	
	Accuracy	CI	Accuracy	CI	Accuracy	CI	Accuracy	CI
LIHC	0.701	0.647	0.742	0.696	0.736	0.681	0.754	0.677
BRCA	0.667	0.653	0.725	0.726	0.717	0.732	0.706	0.715
LUSC	0.648	0.667	0.701	0.727	0.692	0.704	0.689	0.731

4. Discussion

4.1. The effect of the parameter σ in M2DP model.

In the proposed M2DP model, we fix the parameter σ in the weight of the survival loss (*i.e.*, shown in Eq. (2)) as a constant ($\sigma = 1.5$) that is larger than 1. In this section, we investigate the effect of tuning σ in the M2DP model. Specifically, we vary parameter σ from {0.5, 1.5, 2, 2.5}, and record their corresponding Accuracies and Concordance Indexes for both diagnosis and prognosis tasks, respectively. The results are shown in Table 8. As can be seen from Table 8, on one hand, the prediction performance are significantly decreased if $\sigma = 0.5$. The reason lies in that setting σ smaller than 1 overlooks the fact that the genuine survival time for a censored patient is larger than the record survival time for this patient. On the other hand, we can clearly see that the performance of M2DP slightly fluctuates within a very small range if σ is larger than 1, demonstrating that our method is very robust to the choice of σ. Finally, we also find it that the CI for BRCA dataset is lower than that for LUSC cohort, but its F1-score shown in Table 4 is higher than other cancer cohorts. This is because in comparison with the F1 score that is calculated according to the exact stage information in our study, the calculation of CI is based on the patients' survival time that are mostly censored. Intuitively, the ratio of censored patients could affect the CI values since it determines the number of comparable pairs of patients (shown in Eq. (9)). As shown in Table 1, the BRCA dataset has the highest ratio of censored patients, which implies that less comparable pairs are available to calculate the CI. Accordingly,the CI value for BRCA dataset could be lower than the other cohorts although its F1 score can achieve the best.

5. Conclusion

In this paper, we develop M2DP, an effective multi-task multi-modal feature selection method that can jointly identify diagnosis and prognosis associated bio-markers from both histopathological image and gene expression data. The main advantage of our approach is its capability of utilizing the inherent correlation within different tasks to guide the feature selection process, which can more accurately diagnose cancer stage and predict the clinical outcome for different types of cancer patients. M2DP is a general framework and can be easily transferred to other types of cancers, which opens up new opportunity in personalized treatment. In the future, we plan to replace the shallow morphological image feature used in this paper with high-level features learned by deep neural networks for further performance improvement.

Declaration of Competing Interest

Dear Editor-in-Chief:

We would like to submit the enclosed manuscript "Multi-task Multi-modal Learning for Joint Diagnosis and Prognosis of Human Cancers", which is invited to submit to Medical Image Analysis as the extension of our MICCAI 2019 paper (i.e., Diagnosis-Guided Multi-modal Feature Selection for Prognosis Prediction of Lung Squamous Cell Carcinoma).

All authors have approved this submission and confirm there are no conflicts of interest with the requested reviewers.

Best regards,

Wei Shao On behalf of co-authors.

CRediT authorship contribution statement

Wei Shao: Conceptualization, Methodology, Writing - original draft. **Tongxin Wang:** Methodology. **Liang Sun:** Methodology. **Tianhan Dong:** Validation. **Zhi Han:** Writing - original draft. **Zhi Huang:** Conceptualization. **Jie Zhang:** Writing - review & editing. **Daoqiang Zhang:** Supervision. **Kun Huang:** Supervision.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61902183, 61876082, 61861130366, 61732006) and National Key R&D Programme of China (Grant Nos. 2018YFC2001600, 2018YFC2001602), the Royal Society-Academy of Medical Sciences Newton Advanced Fellowship (No. NAF\R1\180371), and the IU Precision Health Initiative Program.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2020.101795.

References

Affo, S., Yu, L.X., Schwabe, R.F., 2017. The role of cancer-associated fibroblasts and fibrosis in liver cancer. Ann. Rev. Pathol. 24, 153–186.
Ai, N., Liu, W., Li, Z., Ji, H., Li, B., Yang, G., 2017. High expression of GP73 in primary hepatocellular carcinoma and its function in the assessment of transcatheter arterial chemoembolization. Oncol. Lett. 14, 3953–3958.
Ashton, T., McKenna, W., Kunz-Schughart, L., Higgins, G., 2018. Oxidative phosphorylation as an emerging target in cancer therapy. Clin. Cancer Res. 24, 2482–2490.
Chen, H.C., Kodell, R.L., Cheng, K.F., 2012. Assessment of performance of survival prediction models for cancer prognosis. BMC Med. Res. Methodol. 12, 1–3.
Cheng, J., Mo, X., Wang, X., 2017. Identification of topological features in renal tumor microenvironment associated with patient survival. Bioinformatics 77, 1–7.
Cheng, J., Mo, X., Wang, X., Huang, K., 2017. Identification of topological features in renal tumor microenvironment associated with patient survival. Bioinformatics 34, 1024–1030.
Cheng, J., Zhang, J., Han, Y., 2017. Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis. Cancer Res. 77, 91–100.
Cho, S., Kim, J., Ryu, H., 2004. Identification of single nucleotide polymorphisms in the tumor necrosis factor (TNF) and TNF receptor superfamily in the korean population. Hum. Immunol. 65, 710–718.
Cooperberg, M., Davicioni, E., Crisan, A., 2015. Combined value of validated clinical and genomic risk stratification tools for predicting prostate cancer mortality in a high-risk prostatectomy cohort. Eur. Urol. 67, 326–333.
Coudray, N., Sakellaropoulos, T., Narula, N., Snuderl, M., 2018. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat. Med. 24, 1559–1563.
Denkert, C., Loibl, S., Noske, A., 2010. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. J. Clin. Oncol. 28, 105–113.
Gecer, B., Aksoy, S., Mercan, E., Shapiro, L., 2018. Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks. Pattern Recognit. 84, 345–356.
Huang, Z., Zhan, X., Xiang, S., 2019. SALMON: Survival analysis learning with multi--omics neural networks on breast cancer. Front Genet. 10, 166–176.
Kim, Y., Kaelin, G., 2004. Role of VHL gene mutation in human cancer. J. Clin. Oncol. 22, 4991–5004.

Lerman, M., Minna, J., 2000. The 630-kb lung cancer homozygous deletion region on human chromosome 3p21. 3: identification and evaluation of the resident candidate tumor suppressor genes. Cancer Res. 60, 6116–6133.
Li, Y., Wang, J., Ye, J., 2016. A multi-task learning formulation for survival analysis. In: Proceedings of the SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1715–1724.
Lin, D., Wei, L.J., Ying, Z., 1993. Checking the cox model with cumulative sums of martingale-based residuals. Biometrika 80, 557–572.
Lin, D.Y., Wei, L.J., 1989. The robust inference for the cox proportional hazards model. J. Am. Stat. Assoc. 84, 1074–1078.
Liu, F., Chen, H., Shen, D., 2014. Inter-modality relationship constrained multi–modality multi-task feature selection for alzheimer's disease and mild cognitive impairment identification. Neuroimage 84, 466–475.
Liu, M., Zhang, D., Chen, S., Xue, H., 2015. Joint binary classifier learning for ECOC-based multi-class classification. IEEE Trans. Pattern Anal. Mach. Intell. 38, 2335–2341.
Liu, Y., Zeng, G., 1997. Cancer and innate immune system interactions: translational potentials for cancer immunotherapy. J. Immunother. 35, 299–305.
Liu, Z., Chen, M., Xie, L., Liu, T., 2018. CLCA4 Inhibits cell proliferation and invasion of hepatocellular carcinoma by suppressing epithelial-mesenchymal transition via PI3k/AKT signaling. Aging (Albany NY) 10, 2570–2577.
Mohammadi, R., Hossein, A., Soltanian, H., 2017. Structured and sparse canonical correlation analysis as a brain-wide multi-modal data fusion approach. IEEE Trans. Med. Imag. 36, 1438–1448.
Niazi, M., Khalid, K., 2016. Visually meaningful histopathological features for automatic grading of prostate cancer. IEEE J. Biomed. Health Inform. 21, 1027–1038.
Nir, G., Hor, S., Karimi, D., Fazli, L., 2018. Automatic grading of prostate cancer in digitized histopathology images: learning from multiple experts. Med. Image Anal. 1, 167–180.
Phoulady, H., Dmitry, B., 2016. Nucleus segmentation in histology images with hierarchical multilevel thresholding. Medical Imaging 2016: Digital Pathology. International Society for Optics and Photonics 9791, 111–117.
Robert, T., 1997. The lasso method for variable selection in the cox model. Stat. Med. 16, 385–395.
Scarpa, A., Mantovani, W., Capelli, P., Beghelli, S., 2010. Pancreatic endocrine tumors: improved TNM staging and histopathological grading permit a clinically efficient prognostic stratification of patients. Mod. Pathol. 23, 824–834.
Shao, W., Cheng, J., Sun, L., Han, Z., Feng, Q., Zhang, D., Huang, K., 2018. Ordinal multi-modal feature selection for survival analysis of early-stage renal cancer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 648–656.
Shao, W., Cheng, J., Zhang, D., Huang, K., 2018. Ordinal multi-modal feature selection for survival analysis of early-stage renal cancer. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 648–656.
Shao, W., Wang, T., Huang, Z., 2019. Diagnosis-guided multi-modal feature selection for prognosis prediction of lung squamous cell carcinoma. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 113–121.
Siegel, R., Miller, K.D., Jemal, A., 2016. Cancer statistics 2016. CA Cancer J Clin 66, 7–30.
Sun, D., Li, A., 2018. Integrating genomic data and pathological images to effectively predict breast cancer clinical outcome. Comput. Method. Progr. Biomed. 161, 45–53.
Thommen, D.S., Schreiner, J., Fischer, O.S., 2015. Progression of lung cancer is associated with increased dysfunction of t cells defined by coexpression of multiple inhibitory receptors. Cancer Immunol. Res. 3, 1344–1355.
Veer, V., Dai, L.J., 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530–532.
Wang, P., Li, Y., Reddy, C., 2019. Machine learning for survival analysis: a survey. ACM Comput. Surv. 51, 1–36.
Wei, L.J., Ying, Z.L., Lin, D.Y., 1990. Linear regression analysis of censored survival data based on rank tests. Biometrika 77, 845–851.
Wilhelm, M., Veltman, A., Kovacs, G., Waldman, F.M., 2002. Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer. Cancer Res. 62, 957–960.
Witten, D.M., Tibshirani, R.J., 2009. Extensions of sparse canonical correlation analysis with applications to genomic data. Stat. Appl. Genet. Mol. Biol. 8, 1–27.
Wu, J., Su, W., 2016. Association between tumor-stroma ratio and prognosis in solid tumor patients: a systematic review and meta-analysis. Oncotarget 18, 68954–68965.
Yang, B., Liu, S., Pang, S., 2018. Deep subspace similarity fusion for the prediction of cancer subtypes. In: IEEE International Conference on Bioinformatics and Biomedicine, pp. 566–571.
Yao, J., Huang, J., 2017. Deep correlational learning for survival prediction from multi-modality data. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 406–414.
Yi, N., Tang, Z., Zhang, X., 2018. BhGLM: bayesian hierarchical GLMs and survival models, with applications to genomics and epidemiology. Bioinformatics 35, 1419–1424.
Yuan, Y., Failmezger, H., Rueda, O.M., 2012. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. Sci. Transl. Med. 4, 143–157.
Zhang, B., Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4, 1–45.
Zhang, J., Huang, K., 2016. Normalized lmQCM: an algorithm for detecting weak quasi-clique modules in weighted graph with application in functional gene cluster discovery in cancer. Cancer Inform. 13, 1–10.
Zhang, Y., Yeung, D.Y., 2010. A convex formulation for learning task relationships in multi-task learning. In: International Conference on Uncertainty in Artificial Intelligence, pp. 733–742.
Zhu, X., Yao, J., Huang, J., 2016. Deep convolutional neural network for survival analysis with pathological images. In: Proceeding of IEEE International Conference on Bioinformatics and Biomedicine, pp. 544–547.
Zhu, Y., Qiu, P., Ji, Y., 2014. TCGA-assembler: open-source software for retrieving and processing TCGA data. Nature Methods 11. 599–599.