Methods 189 (2021) 74–85

Contents lists available at ScienceDirect

Methods

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

Integrating multi-omics data by learning modality invariant representations for improved prediction of overall survival of cancer

Li Tong [a], Hang Wu [b], May D. Wang [a,*]

[a] Dept. of Biomedical Engineering, Georgia Tech and Emory University, Atlanta, USA
[b] Dept. of Biomedical Engineering, Georgia Tech, Atlanta, USA

ARTICLE INFO

Keywords:
Multi-omics integration
Survival analysis
Consensus learning
Modality-invariant representation

ABSTRACT

Breast and ovarian cancers are the second and the fifth leading causes of cancer death among women. Predicting the overall survival of breast and ovarian cancer patients can facilitate the therapeutics evaluation and treatment decision making. Multi-scale multi-omics data such as gene expression, DNA methylation, miRNA expression, and copy number variations can provide insights on personalized survival. However, how to effectively integrate multi-omics data remains a challenging task. In this paper, we develop multi-omics integration methods to improve the prediction of overall survival for breast cancer and ovarian cancer patients. Because multi-omics data for the same patient jointly impact the survival of cancer patients, features from different -omics modality are related and can be modeled by either association or causal relationship (e.g., pathways). By extracting these relationships among modalities, we can get rid of the irrelevant information from high-throughput multi-omics data. However, it is infeasible to use the Brute Force method to capture all possible multi-omics interactions. Thus, we use deep neural networks with novel divergence-based consensus regularization to capture multi-omics interactions implicitly by extracting modality-invariant representations. In comparing the concatenation-based integration networks with our new divergence-based consensus networks, the breast cancer overall survival C-index is improved from 0.655 ± 0.062 to 0.671 ± 0.046 when combing DNA methylation and miRNA expression, and from 0.627 ± 0.062 to 0.667 ± 0.073 when combing miRNA expression and copy number variations. In summary, our novel deep consensus neural network has successfully improved the prediction of overall survival for breast cancer and ovarian cancer patients by implicitly learning the multi-omics interactions.

1. Introduction

The advancement of biomedical techniques such as next-generation sequencing (NGS) and wearable devices have generated high-throughput multi-modality data enabling a more comprehensive view of the patients for personalized and precision care. Besides lab tests and medical imaging, physicians can order genetic tests to obtain patient genomics information to make more accurate diagnoses and decisions. However, compared to traditional features such as vital signs and lab tests, it is infeasible for physicians to identify useful patterns directly from the high-throughput genomics data with millions of features.

High-throughput multi-omics data have enabled personalized diagnosis and treatments for genetics related diseases such as cancer. However, it is extremely challenging for physicians to make sense of the molecular biomarkers directly from the -omics data tsunami. For computational scientists to integrally analyze multi-omics data, there are also a few issues to be addressed:

1) "the curse of dimensionality" or "big-p-small-n" issue [1]: the number of samples n typically ranges from 100 to 1000, but the number of -omics features p is much larger, with can be up to millions. This "big-p-small-n" problem often leads to model over-fitting in practice. Thus, we need to design specific feature selection or extraction methods to eliminate unrelated features and reduce the number of features used for downstream tasks.
2) Multi-modal data analysis, including feature selection and predictive modeling, is more complex than single-modality analysis due to the cross-modality information. There are three ways to address this: to consider the cross-modality interaction and identify correlated features among modalities in feature selection stage [2]; to integrate the multi-modality information at the intermediate feature stage [3–5]; or to integrate at the decision stage [6]. Although

* Corresponding author.
E-mail addresses: ltong@gatech.edu (L. Tong), hangwu@gatech.edu (H. Wu), maywang@gatech.edu (M.D. Wang).

https://doi.org/10.1016/j.ymeth.2020.07.008
Received 16 January 2020; Received in revised form 29 June 2020; Accepted 21 July 2020
Available online 5 August 2020

substantial progress has been made in this area, modeling the interactions among modalities and getting rid of the redundant or irrelevant information remains extremely challenging.

Data integration algorithm investigates on how to combine multi-modal features of the same patient to achieve improved prediction performance. In multi-omics data integration, we aim to improve the diagnosis (e.g., cancer staging) and prognosis prediction (e.g., survival analysis) by combing the information embedded in each modality. We categorize the multi-omics integration methods based on whether a specific class of computational models is used. Model-agnostic approaches integrate features from each modality directly, either by concatenating them after feature selection [7,2] or integrating them after decision made of each modality [6]. In model-based approaches, models such as kernel machines, graphical models, and neural networks [3] are needed to encode the assumptions of data. Then based on the designed models, features are extracted for downstream tasks. Because multi-omics variations jointly impact disease development and patient prognosis, the good data integration methods can improve the model performance by capturing the interactions among multi-omics (e.g., gene expression pathways) compared to single-omics methods. However, many multi-omics integration studies in the literature can hardly exploit the full interactions among modalities and thus result in sub-optimal performance. Because it is difficult to explore all multi-omics interactions explicitly, we investigate how to learn the interactions implicitly by deep neural networks with divergence-based regularization for classification and survival modeling tasks. Our novel contributions include three parts:

1. We develop an effective end-to-end solution for integrating different modalities based on deep neural networks. We develop a novel regularization technique to model the observation that two modalities share consensus information about the same patient. This method can be easily extended to multiple modalities.
2. We present evidence and suggest when and how data integration should be done to integrate different modalities of data.
3. We show the effectiveness of the new divergence-based consensus regularization by extensive experiments on cancer types classification and cancer survival prediction, and by visualization of the modality-invariant features after consensus learning.

2. Related works

2.1. Multi-view/multi-modal learning

Multi-view/ multi-modal learning [8,9] are machine learning techniques for building models that can integrate multiple types of input information (called 'views' or 'modalities' in the literature). For example, in the biomedical field, a patient's electronic health records (EHRs), personal health records (PHRs), genomic and proteomic variations, and medical images can be considered various views. Ideally, these multi-view data should be jointly evaluated to realize a more personalized diagnosis and treatment.

The most naive approach is to treat the multi-view learning problem as a single-view problem, where all views are concatenated into a single view and then solved with the established single-view models. However, as the feature concatenation ignores the interactions between views and the number of features increases as the number of modalities increases, the performance of this naive multi-view method can be sub-optimal with issues such as over-fitting.

To exploit the multi-view data, we have to consider the interactions among modalities. There are two principles in multi-view learning [8]: 1) the consensus principle, which assumes that the disagreement between views upper bounds the classification errors. Thus, we should aim to maximize the agreement between views. 2) the complementary principle, which assumes that each view contains information other views do not have, and we should extract the difference from each view while preserving the shared information. Based on these two principles, researchers have developed two approaches, model-agnostic and model-based, to combine data from multiple modalities.

Model-agnostic approaches are simple in design and usually utilize the complementary principle. Based on when data from different modalities are integrated, we have early integration (where we concatenate the raw/pre-processed features), late integration (where we combine the output from each learning algorithm), and hybrid integration (where we use early and late integration together).

Model-based approaches design models for integrating different modalities: 1) kernel-based algorithms, mainly multiple kernel learning (MKL) [10], first compute kernel matrices for each modality, then combines kernels in a linear or non-linear fashion for succeeding kernel-based classification or regression algorithms. As kernels evaluate the similarities between data points, using modality-specific kernels helps capture heterogeneous information from each modal and improves the performance. Kernel methods are especially helpful for small sample sizes but suffer from high computational complexity when sample sizes are large. 2) graphical models such as Bayesian networks and Markov random fields are a class of algorithms that treat each feature as a random variable and exploit the probability relationships among them [11]. The benefit of this approach is that we can incorporate more priors into our modeling and easily interpret the models. However, graphical models are also computationally expensive. 3) deep learning-based multi-modal learning algorithms [12] gain increasing popularity in the literature for the past few years. For each modality, we design a modality-specific neural network for feature extraction, and the extracted features can be fused for downstream analysis. The benefit of this approach is that deep neural networks excel in extracting non-linear features and can easily incorporate additional regularization. Popular architectures for deep learning-based multi-modal integration include joint representation, coordinated representation, and cross-modality autoencoders [13]. The whole architecture can be trained end-to-end with gradient-based optimization algorithms, making the approach scalable to large sample scenarios.

2.2. Multi-modal learning in biomedical science

With the development of data collection technologies in the biomedical domain, researchers now have access to various multi-modal data such as high-throughput multi-omics data (e.g., gene expression, DNA methylation, and single nucleotide polymorphisms (SNPs)), medical imaging data (e.g., pathology and radiology), and clinical records (e.g., demographics, insurance claims, and past medication history) [14,15].

Based on characteristics of the modality available, researches have designed various task-specific multi-modal learning algorithms in recent biomedical research [16–19]. In the sub-filed of -omics data analysis, various multi-modal learning methods have been proposed for multi-omics integration. For example, Xu et al. [20] studied how to integrate-omics data using graph-based similarity between molecular information such as gene expression, DNA methylation, and showed superior performance in cancer types classification and survival prediction. Vasaikar et al. [21] created a database of over one billion data points by combining multi-omics data and clinical data from the TCGA dataset for 32 cancer types with the proteomics data. In addition, they presented a module for analyzing and visualizing associations between clinical and molecular attributes. Way et al. [22] integrated RNA-seq, copy number variations, and mutations for identification of abnormal molecular states in tumors. Ma et al. [4] studied how to integrate multi-omics data using neural network-based approaches by combining multiple autoencoders and how domain knowledge can be incorporated to improve the learned representations. The improved representations outperform other integration methods in predicting disease progression on TCGA datasets. A similar autoencoder-based approach has been

proposed by Chaudhary et al. [7], where they used k-means clustering to identify survival-risk subgroups and showed that integration of mRNA, miRNA, and DNA methylation improved the survival prediction for cancer patients. Huang et al. [5] integrated intermediate representations from multi-layer perceptrons and showed improved performance on cancer survival prediction on select TCGA datasets.

In this paper, we propose to integrate multi-omics data with consensus learning (Fig. 1). To implicitly model the interactions among modalities, we focus on maximizing the agreement between modalities. By learning modality-invariant representations with divergence-based consensus regularization, we integrate the multi-omics data in a common hidden space and improve the prediction of overall survival for breast cancer and ovarian cancer patients.

3. Materials and methods

3.1. Datasets

For cancer types classification, we collect four TCGA cancer datasets from UCSC Xena [23] including lung adenocarcinoma (LUAD), kidney renal clear cell carcinoma (KIRC), lung squamous cell carcinoma (LUSC), and pancreatic adenocarcinoma (PAAD) (Table 1). Each cancer dataset consists of four -omics data with the same numbers of features, including gene expression (GeneExp), DNA methylation (DnaMeth), miRNA expression (miRNA), and copy number variations (CNVs). The GeneExp and miRNA data downloaded from UCSC Xena has already been log2 transformed. We apply min-max transformation to all four -omics data and normalize features to the range $(0,1)$. For each cancer type, we only keep the samples with all four -omics data for the cancer types classification. For every classification experiment configuration, we choose two out of the four modalities as the input, and we exhaust all six combinations for our experiments.

For overall survival analysis, we collect two cancer datasets directly from the TCGA data portal, including breast cancer (BRCA) and ovarian cancer (OV). Each cancer dataset also consists of the GeneExp, DnaMeth, miRNA, and CNVs data along with overall survival information (Table 2). The number of features is 60,483 for GeneExp, 25,978 for DnaMeth, 1,881 for miRNA, and 19,729 for CNVs in overall survival analysis. Note that we use different DnaMeth experiment data as that in the cancer types classification. We first apply log2 transformation to GeneExp and miRNA data and then perform min-max normalization to scale the features to the range (0, 1) for all -omics features. Similarly, we only keep samples with all four -omics data for overall survival analysis.

Table 1
Overview of the TCGA samples for cancer types classification.

Cancer Types	GeneExp	DneMeth	miRNA	CNVs	# of Samples with All Modalities
LUAD	585	503	564	531	454
KIRC	607	483	592	536	319
LUSC	550	412	523	503	364
PAAD	182	195	183	185	177
# of features	60,483	485,577	1,881	19,729	

3.1.1. Feature Selection and dimension reduction

A typical -omics data set usually contains only a few hundred samples but with millions of features. Thus, -omics data usually suffer from the "curse of dimensionality." Researchers typically apply various feature selection or dimension reduction techniques to the raw- omics features to mitigate the challenge. For example, Huang et al. apply co-expression analysis to reduce the number of gene expression and miRNA expression features [5]. EL-Manzalawy et al. utilize min-redundancy and max-relevance for feature selection for multi-omics data [2]. In this study, we focus on integration methods and apply two simple feature selection or dimension reduction techniques. The first method we choose is the dimension reduction technique principal component analysis (PCA). We apply PCA to each -omics data and obtain the first n principal components(PCs) as the dimension-reduced features for integration. We will call these features as PCA features hereafter. To determine the optimal numbers of PCs to keep, we apply a grid search for each task based on the sample sizes, as the number of PCs cannot exceed the training sample sizes. For cancer types classification, we apply the grid search for PCA features with $n = 50,100,150,200$. We do not include the number of PCs beyond 200 as the performance improvement is marginal for cancer types classification. For breast cancer overall survival prediction, we apply the grid search for PCA features with n ranges from 50 to 600 with a step size of 50. For ovarian cancer overall survival prediction, we apply the grid search for PCA features with $n = 50,100,150,200$, as the sample size of ovarian cancer is smaller than that of breast cancer.

The second method we choose is the unsupervised univariate feature selection by variance. For each -omics data, we choose the top 1000 features with the largest variances. We will call these features high variance features hereafter. For simplicity, we do not apply gird search for high variance features to determine the optimal number of features as we did for PCA features. The baseline models and proposed integration models are applied to both PCA features and high variance features.

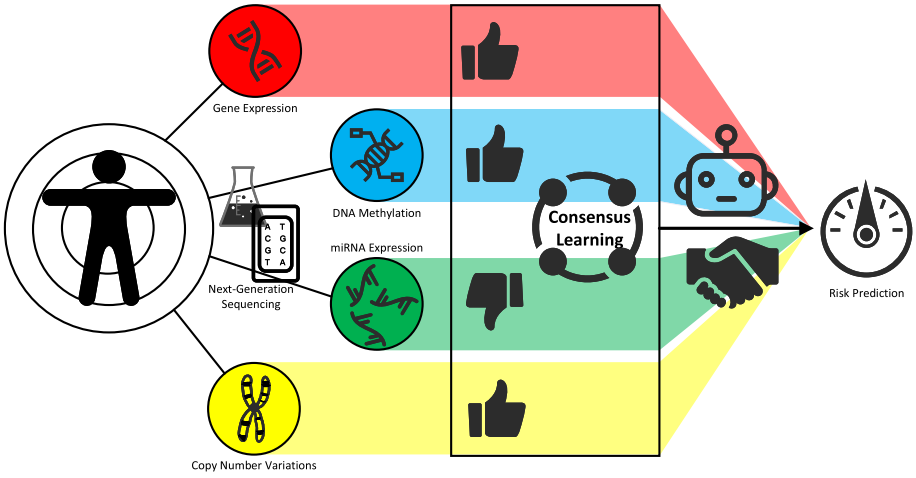

Fig. 1. Integration of multi-omics data (e.g., gene expression, DNA methylation, miRNA expression, and copy number variations (CNVs)) with consensus learning for improved prediction performance.

Table 2
Overview of the TCGA samples for overall survival analysis.

Cancer Types	GeneExp	DneMeth	miRNA	CNVs	# of Samples with All Modalities	# of Event	Max Survival Time	Min Survival Time
BRCA	1,222	1,234	1,207	1,106	1,061	145	8605	0
OV	379	613	499	620	362	221	5481	8

3.1.2. Train test split

For both cancer types classification and overall survival analysis experiments, we perform a stratified fourfold cross-validation with 60% training, 15% validation, and 25% testing data. The cancer types classification experiments are stratified with cancer types as the label, and the overall survival analysis experiments are stratified with survival events.

3.2. Consensus feature representation learning for multi-modality data integration

In this study, we propose to integrate multi-omics data with consensus constraints, aiming to generate modality-invariant representations among various -omics data. We compare the performance of the proposed network architecture with the baselines, including the single-

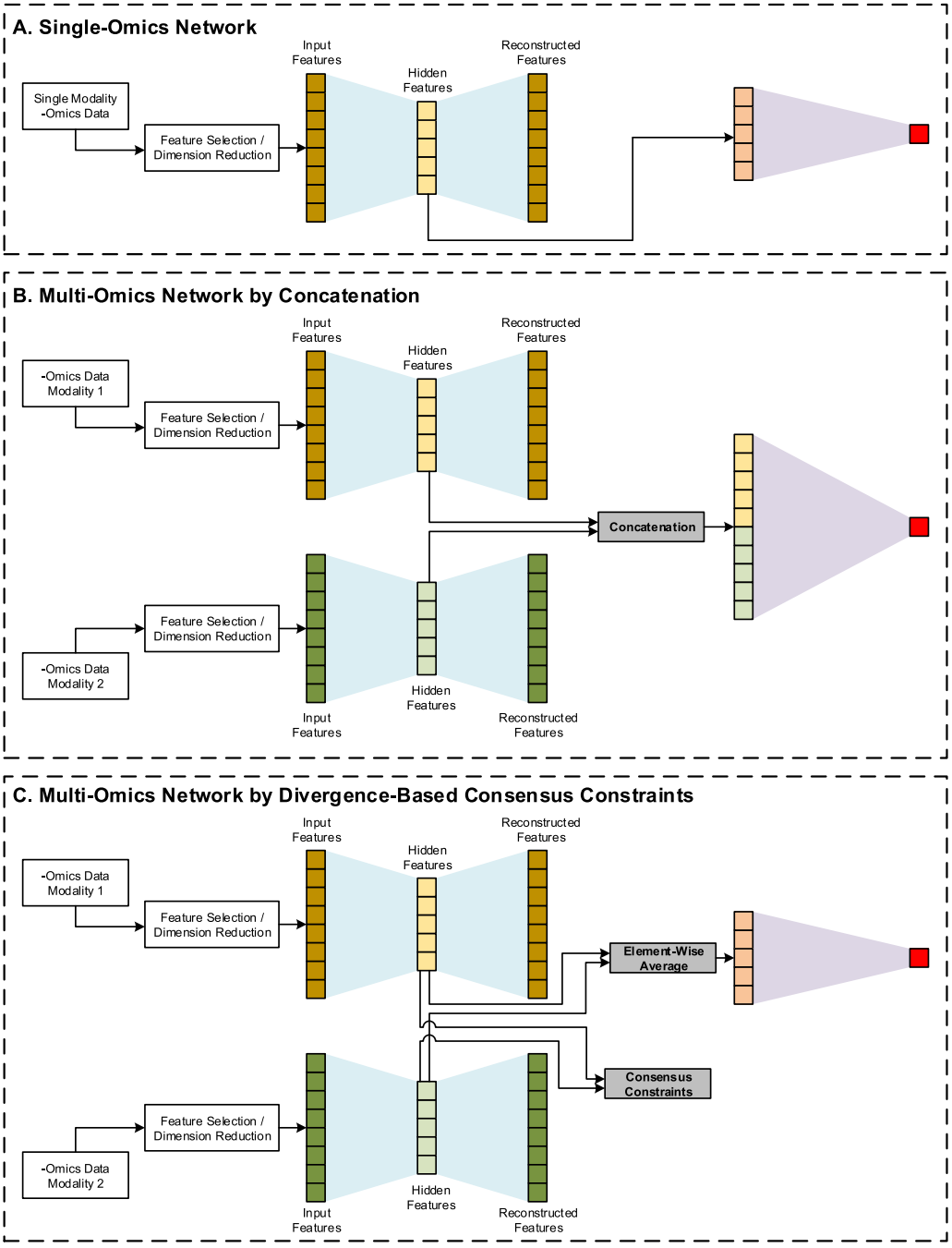

Fig. 2. The network architectures for single-omics and multi-omics integration. A. Single-omics netowrk. B. Multi-omics network by concatenation. C. Multi-omics network by divergence-based consensus constraints.

modality network and concatenation-based integration network with or without the autoencoding process. The network architectures compared in this study are visualized in Fig. 2. In the single-omics network (Fig. 2A), we use an encoder for feature representation and a decoder for reconstruction. The represented hidden features are fed into a task-specific network for classification or survival analysis.

In the concatenation-based multi-omics integration network (Fig. 2B), we have an encoder-decoder structure for each -omics modality, the represented hidden features are concatenated and then fed into the task-specific network (Algorithm 1). We will call this network as AutoencoderConcat. We also test the performance using the concatenation-based network without the decoder part [5], which is called as SimpleConcat.

Inspired by the integrative analysis of -omics data by dimension reduction to correlated structure across modalities [24], we propose to integrate the multi-omics data with consensus constraints (i.e., domain-invariant representations) as shown in Fig. 2C. With the encoder-decoder framework for each -omics modality, we impose a divergence-based constraint on the hidden features learned from each modality (Algorithm 2). In this study, we tested the consensus learning with the Cosine similarity [4] and Euclidean distance, respectively. For consensus learning with Cosine similarity, we maximize the Cosine similarity between hidden features learned from each -omics modality. We will refer to this framework as ConsensusCosine. We minimize the Euclidean distance between hidden features learned from each -omics modality for consensus learning with Euclidean distance. We will refer to this framework as ConsensusEuclidean. The learned hidden features from concatenation-based integration and consensus-based integration are visualized with the t-SNE plot for comparison.

For single-modality networks, concatenation-based integration networks (SimpleConcat and AutoencoderConcat), and consensus-based integration networks (ConsensusCosine and ConsensusEuclidean), we utilize the same network structures for the shared components. For simplicity, we have used three-layer fully connected networks with various numbers of neurons for encoder, decoder, and task-specific networks, respectively.

Algorithm 1 Training Concatenation-Based Classification Models

Require: Multi-modal Dataset $D = \{(X_i = (M_{i,1}, M_{i,2}), y_i)\}$
Require: $Euc(a,b)$ as the Euclidean distance between vectors a, b
  Initialize modality specific encoders as Enc_1, Enc_2
  Initialize modality specific encoders as Dec_1, Dec_2
  Initialize classification module as CLF
  whileNot Converged
    //**Encoder-Decoder Training**
    Sample a batch of sample from D
    $loss \leftarrow 0$.
    for each sample i in the batch **do**
      Compute reconstructed input as $\widehat{M}_{i,j} = Dec_j(Enc_j(M_{i,j}))$, for $j = 1, 2$
      Compute reconstruction loss $recon\_loss_i$ as $Euc(\widehat{M}_{i,1}, M_{i,2}) + Euc(\widehat{M}_{i,2}, M_{i,2})$
      $loss \leftarrow loss + recon\_loss_i$
    end for
    Back propagate $loss$ and update model parameters for encoders and classification module
    //**Encoder-Classifier Training**
    Sample a batch of sample from D
    $loss \leftarrow 0$.
    for each sample i in the batch **do**
      Compute modality specific representations as $h_{i,j} = Enc_1(M_{i,j})$, for $j = 1, 2$
      Build joint representation as $h_i = Concat(h_{i,1}, h_{i,2})$
      Compute classification loss clf_loss_i using y_i and prediction $\widehat{y} = CLF(h_i)$
      $loss \leftarrow loss + clf\_loss_i + reg_i$
    end for
  Back propagate $loss$ and update model parameters for encoders and classification module
  end while
  return Modality specific encoders and decoders $Enc_1, Enc_2, Dec_1, Dec_2$, and classification module as CLF

Algorithm 2 Training Consensus-Based Classification Models

Require: Multi-modal Dataset $D = \{(X_i = (M_{i,1}, M_{i,2}), y_i)\}$
Require: $Euc(a,b)$ as the Euclidean distance between vectors a, b
  Initialize modality specific encoders as Enc_1, Enc_2
  Initialize modality specific encoders as Dec_1, Dec_2
  Initialize classification module as CLF
  while Not Converged
    //**Encoder-Decoder Training**
    Sample a batch of sample from D
    $loss \leftarrow 0$.
    for each sample i in the batch **do**
      Compute reconstructed input as $\widehat{M}_{i,j} = Dec_j(Enc_j(M_{i,j}))$, for $j = 1, 2$
      Compute reconstruction loss $recon\_loss_i$ as $Euc(\widehat{M}_{i,1}, M_{i,2}) + Euc(\widehat{M}_{i,2}, M_{i,2})$
      $loss \leftarrow loss + recon\_loss_i$
    end for
    Back propagate $loss$ and update model parameters for encoders and classification module
    //**Encoder-Classifier Training**
    Sample a batch of sample from D
    $loss \leftarrow 0$.
    for each sample i in the batch **do**
      Compute modality specific representations as $h_{i,j} = Enc_1(M_{i,j})$, for $j = 1, 2$
      Build joint representation as $h_i = \frac{1}{2}(h_{i,1} + h_{i,2})$
      Compute classification loss clf_loss_i using y_i and prediction $\widehat{y} = CLF(h_i)$
      Compute consensus regularization, for example, the Euclidean-based regularization as $reg_i = Euc(h_{i,1}, h_{i,2})$
      $loss \leftarrow loss + clf\_loss_i + reg_i$
    end for
    Back propagate $loss$ and update model parameters for encoders and classification module
  end while
  return Modality specific encoders and decoders $Enc_1, Enc_2, Dec_1, Dec_2$, and classification module as CLF

3.3. Endpoint 1: cancer types classification

The first endpoint for our proposed multi-omics integration network is multi-class classification. For this endpoint, we use a fully-connected network for classification and trained with cross-entropy loss. The classification performance is evaluated by accuracy, precision, recall, and area under the curve (AUC) for binary classification. For multi-class classification, we use accuracy, weighted precision, and weighted recall as the evaluation metrics. These metrics are in the range of $[0,1]$, and the higher, the better.

3.4. Endpoint 2: survival risk analysis

Survival analysis aims to predict the expected duration of time until one or more events happen by modeling the time to event data. The proportional hazards model assumes the covariates are multiplicatively related to the hazard [25]. Assuming the proportional hazards assumption holds, the Cox proportional hazards model can estimate the effect parameters without considering the hazard function [26]. Thus, the Cox proportional hazards model is semi-parametric. Let $X_i = X_{i1}, \ldots, X_{ip}$ be covariates for subject i. The hazard function for the Cox proportional hazards model has the form:

$$\lambda(t|X_i) = \lambda_0(t)\exp(\beta_1 X_{i1} + \ldots + \beta_p X_{ip}) = \lambda_0(t)\exp(X_i \cdot \beta) \tag{1}$$

This expression gives hazards function at time t for subject i with covariate vector X_i. The likelihood of the event to be observed occurring for subject i at time Y_i can be written as:

$$L_i\left(\beta\right) = \frac{\lambda(Y_i|X_i)}{\sum\limits_{j:Y_j \geqslant Y_i} \lambda\left(Y_i \middle| X_j\right)} = \frac{\lambda_0(Y_i)\theta_i}{\sum\limits_{j:Y_j \geqslant Y_i} \lambda_0\left(Y_i\right)\theta_j} = \frac{\theta_i}{\sum\limits_{j:Y_j \geqslant Y_i} \theta_j} \tag{2}$$

where $\theta_j = \exp(X_j \cdot \beta)$ and the summation is over the set of subjects j where the event has not occurred before time Y_i (including subject i itself). $L_i(\beta)$ is called a partial likelihood as the effect of the joint probability can be estimated without modeling the change of the hazard over time. Obviously $0 < L_i(\beta) < 1$. Treating the subjects as if they were statistically independent of each other, we can obtain the joint probability of all realized events:

$$L\left(\beta\right) = \prod_{i:C_i=1} L_i\left(\beta\right) \tag{3}$$

where the occurrence of the event is indicated by $C_i = 1$. The corresponding log partial likelihood is

$$l\left(\beta\right) = \sum_{i:C_i=1}\left(X_i \cdot \beta - \log \sum_{j:Y_j \geq Y_i} \theta_j\right) \tag{4}$$

With the development of deep learning, the Cox proportional hazards model has been extended with deep neural networks. Cox-Time [27] and Deep Surv [28] similarly replacing the linear relationship $\exp(X_i \cdot \beta)$ with non-linear transformation as $\exp(f_\phi(X_i) \cdot \beta)$, where f_ϕ is a neural network parametrized by ϕ, for example, a fully connected neural network. In addition, the authors also proposed L_1 and L_2 regularization terms, respectively, on the parameter ϕ to reduce over-fitting of the models. In this study, we use a fully connected neural network for the Cox proportional hazards model with the log partial likelihood as the loss function.

To evaluate the risk scores produced by survival models, researchers have developed various metrics to measure the concordance between the predicted risk scores and the actual survival time. Following the previous studies in deep-learning-based survival analysis [5], we evaluate the overall survival analysis performance with the concordance index (C-index) [29]. C-index evaluates how well the survival risk we computed aligns with the actual survival time, i.e., for two individuals $(X_1, T_1), (X_2, T_2)$, $\mathrm{C-index} = \Pr(\lambda(X_2) < \lambda(X_1) | T_2 > T_1)$. We also evaluate the survival analysis with the Kaplan-Meier curve and the log-rank test.

3.5. Implementation and experiments

The train-test split for cross-validation, the classification metrics, the t-SNE visualization are implemented with [30]. The Kaplan-Meier plot and log-rank test are performed with lifelines [31]. The neural networks are designed and implemented with PyTorch 1.1.0. For cancer types classification, we use a batch size of 8, Adam optimizer with a learning rate of 0.001, and training epochs of 200. The test performance is based on the models with the best validation performances. For survival analysis, we use a batch size of 128, Adam optimizer with a learning rate of 0.001, and training epochs of 200. As we observed an over-fitting for models with the best validation performance, the test performance is based on the models until the end of training (200 epochs). More details of the model implementation and training details can be found at Github repo (https://github.com/tongli1210/TCGA_Omics).

4. Results

4.1. Cancer types classification

We perform cancer types classification using four TCGA datasets (LUAD, KIRC, LUSC, and PAAD) with six binary classifications and one four-class classification. The binary classification is evaluated with accuracy, precision, recall, and AUC. The four-class classification is evaluated with accuracy, precision, and recall. For each cancer-types classification task, we perform PCA feature reduction with a grid search of the number of PCs from 50 to 200 with a step size of 50. The percentage of variances explained by the PCs are presented in Supplementary Material S1.1 for each classification task and train-test split. Based on the results, we observe that the first 50 PCs can explain about 70%, 90%, 80%, 70% of variances for GeneExp, DnaMeth, miRNA, and CNV respectively.

We perform cancer types classification using two of the proposed frameworks AutoencoderConcat and ConsensusEuclidean. The classification results using various numbers of PCA features and high variance features are presented in Supplementary Materials S2. Furthermore, we visualize the accuracy of integration-based cancer types classification for some selected experiments using radar plots in Fig. 3 and Supplementary Materials S4. For single-omics binary classification, GeneExp, DnaMeth, and miRNA achieve similar performance and significantly outperform CNV (Supplementary Materials S2). For multi-omics binary classification, the accuracy for both integration frameworks AutoencoderConcat or ConsensusEuclidean is very high except for LUAD vs. LUSC. Since LUAD and LUSC are both lung cancers, we suspect the separation of LUAD vs. LUSC is harder than the other binary classification tasks, which also explains the performance drop for four-class classification that includes both LUAD and LUSC. Comparing the two multi-omics integration approaches, we find that the AutoencoderConcat consistently outperforms the ConsensusEuclidean integration when using PCA features. In contrast, the AutoencoderConcat integration consistently outperforms the ConsensusEuclidean integration when using high variance features (Fig. 3).

4.2. Cancer overall survival analysis

4.2.1. Evaluation of overall survival analysis using C-index

Similar to the cancer types classification task, we perform PCA feature reduction with a grid search of the number of PCs for breast cancer and ovarian cancer overall survival prediction, respectively. For breast cancer, the percentages of variances explained by the PCs are presented in Supplementary Material S1.2 for various modalities and train-test splits. Based on the results, we observe that the first 50 PCs can explain about 72%, 92%, 82%, 64% of variances for GeneExp, DnaMeth, miRNA, and CNV respectively. For ovarian cancer, the percentages of variances explained by the PCs are presented in Supplementary Material S1.3 for various modalities and train-test splits. Based on the results, we observe that the first 50 PCs can explain about 72%, 91%, 84%, 71% of variances for GeneExp, DnaMeth, miRNA, and CNV respectively.

We evaluate the overall survival analysis performance with PCA features and high variance features using C-index for both single-omics and multi-omics approaches (Supplementary Material S3). For breast cancer (BRCA) survival analysis, the best performance for PCA features is achieved with 300 PCs. Thus, we select the results using 300 PCs in Table 3 for single-omics experiments and Table 4 for multi-omics experiments. For ovarian cancer (OV) survival analysis, the best performance for PCA features is achieved with 50 PCs. Thus, we select the results using 50 PCs in Table 5 for single-omics experiments and Table 6 for multi-omics experiments.

From the tables, we can observe that the performance of overall survival analysis improves after the integration of multi-omics with both PCA features or high variance features. For BRCA survival analysis, the concatenation-based integration outperforms consensus-based integration in some -omics combinations while consensus-based integration outperforms concatenation-based integration in the other -omics combinations. The consensus-based integration ConsensusCosine using PCA features of DnaMeth and miRNA achieves the best performance (0.671 ± 0.046). The consensus-based integration ConsensusEuclidean using PCA features of miRNA and CNV achieves a similar performance (0.667 ± 0.073).

For OV survival analysis, we also observed that the concatenation-based integration outperforms consensus-based integration in some -omics combinations while consensus-based integration outperforms concatenation-based integration in the other -omics combinations. The concatenation-based integration AutoencoderConcat using PCA features of miRNA and CNV achieves the best performance (0.571 ± 0.036). We

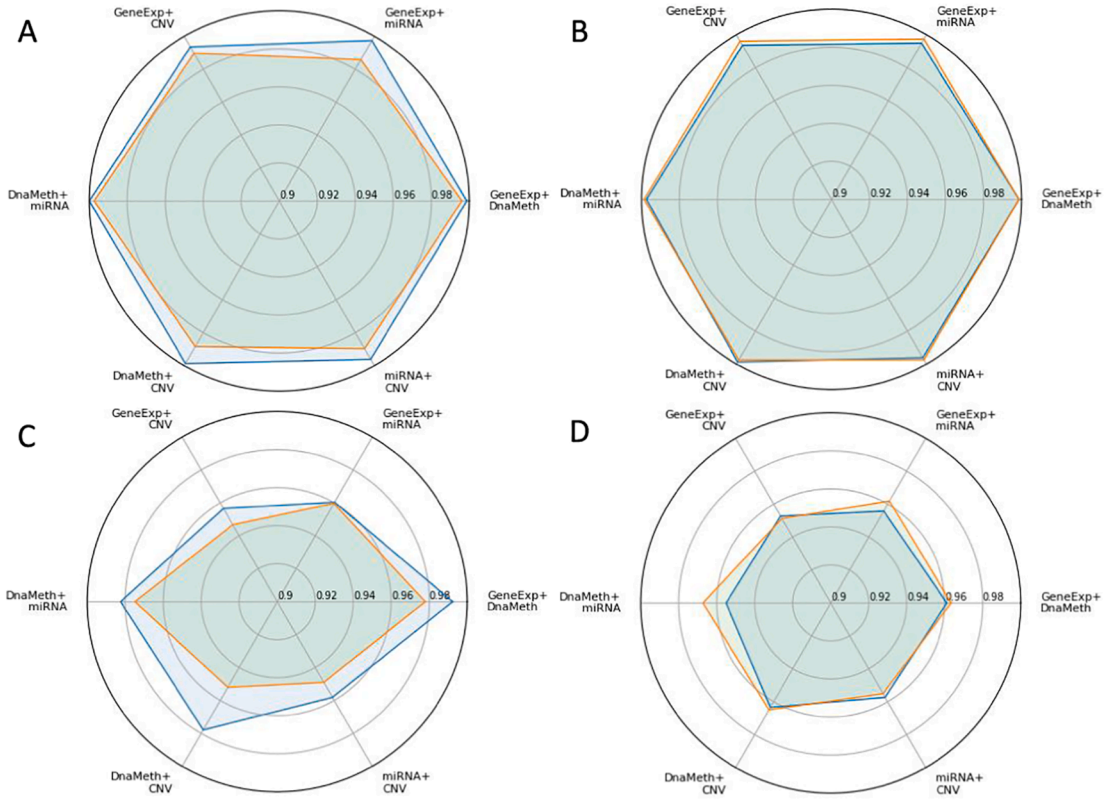

Fig. 3. Radar plots of the accuracy for cancer types classification using AutoencoderConcat framework and ConsensusEuclidean framework for integration. A. LUAD vs. KIRC binary classification using the top 100 PCA features. B. LUAD vs. KIRC binary classification using the high variance features. C. LUAD vs. KIRC vs. LUSC vs. PAAD four-class classification using the top PCA features. D. LUAD vs. KIRC vs. LUSC vs. PAAD four-class classification using the high variance features. Blue lines: AutoencoderConcat. Yellow lines: ConsensusEuclidean.

Table 3
BRCA Overall Survival (OS) Analysis C-Index with Single-Omics.

Features	GE	DM	mR	CNV
PCA 300	0.585 ± 0.065	0.591 ± 0.064	0.629 ± 0.089	0.496 ± 0.051
High Variance	0.529 ± 0.033	0.581 ± 0.066	0.614 ± 0.041	0.503 ± 0.071

have also compared the C-index of concatenation-based and consensus-based integration methods using radar plot (Fig. 4).

4.2.2. Visualization of risk prediction with Kaplan-Meier plot

Besides the C-index for the evaluation of survival analysis, we also perform the Kaplan-Meier plot with the log-rank test to visualize the hazard prediction. We group the testing samples into the low-risk and high-risk groups by the median of all predicted hazards. If a testing sample's predicted hazard is lower than the median hazard of all testing samples, it will be assigned to the low-risk group; otherwise, it will be assigned to the high-risk group. We visualize the two groups with Kaplan-Meier curves and test the separation of these two groups with the log-rank test. Fig. 5 presents the Kaplan-Meier plot of breast cancer (BRCA) overall survival (OS) analysis using the integration of the top 300 PCA features of DnaMeth and miRNA. From this Kaplan-Meier plot, we observe that all four integration methods are able to achieve good separation of the low-risk and high-risk groups. Fig. 6 presents the Kaplan-Meier plot of breast cancer (BRCA) overall survival (OS) analysis

Table 5
OV Overall Survival (OS) Analysis C-Index with Single-Omics.

Features	GE	DM	mR	CNV
PCA 50	0.504 ± 0.016	0.488 ± 0.046	0.549 ± 0.037	0.523 ± 0.037
High Variance	0.49 ± 0.024	0.524 ± 0.026	0.506 ± 0.046	0.497 ± 0.03

Table 4
BRCA Overall Survival (OS) Analysis C-Index with Multi-Omics.

Features	Methods	GE + DM	GE + mR	GE + CNV	DM + mR	DM + CNV	mR + CNV
PCA 300	SimpleConcat	0.596 ± 0.054	**0.644 ± 0.057**	0.583 ± 0.066	0.651 ± 0.055	0.562 ± 0.057	0.608 ± 0.039
	AutoencoderConcat	**0.607 ± 0.073**	0.609 ± 0.053	**0.588 ± 0.105**	0.655 ± 0.062	0.552 ± 0.061	0.627 ± 0.062
	ConsensusCosine	0.558 ± 0.081	0.609 ± 0.078	0.571 ± 0.054	**0.671 ± 0.046**	0.565 ± 0.024	0.609 ± 0.056
	ConsensusEuclidean	0.581 ± 0.057	0.591 ± 0.115	0.52 ± 0.025	0.637 ± 0.073	**0.618 ± 0.086**	**0.667 ± 0.073**
High Variance	SimpleConcat	0.524 ± 0.024	0.525 ± 0.04	**0.558 ± 0.019**	**0.633 ± 0.042**	0.577 ± 0.04	0.609 ± 0.054
	AutoencoderConcat	0.507 ± 0.036	0.53 ± 0.052	0.524 ± 0.038	0.625 ± 0.023	**0.586 ± 0.068**	0.603 ± 0.04
	ConsensusCosine	**0.532 ± 0.017**	0.543 ± 0.011	0.494 ± 0.052	0.61 ± 0.068	0.557 ± 0.034	**0.61 ± 0.056**
	ConsensusEuclidean	0.528 ± 0.018	**0.606 ± 0.041**	0.532 ± 0.02	0.626 ± 0.056	0.554 ± 0.024	0.585 ± 0.049

Table 6
OV Overall Survival (OS) Analysis C-Index with Multi-Omics.

Features	Methods	GE + DM	GE + mR	GE + CNV	DM + mR	DM + CNV	mR + CNV
PCA 50	SimpleConcat	**0.519 ± 0.034**	0.505 ± 0.033	0.548 ± 0.015	**0.529 ± 0.075**	0.517 ± 0.044	0.545 ± 0.057
	AutoencoderConcat	0.462 ± 0.027	**0.556 ± 0.059**	0.503 ± 0.027	0.513 ± 0.014	0.489 ± 0.074	**0.571 ± 0.036**
	ConsensusCosine	0.497 ± 0.011	0.507 ± 0.043	**0.553 ± 0.053**	0.497 ± 0.053	**0.535 ± 0.037**	0.533 ± 0.036
	ConsensusEuclidean	0.496 ± 0.039	0.548 ± 0.025	0.512 ± 0.026	0.485 ± 0.047	0.508 ± 0.026	0.54 ± 0.067
High Variance	SimpleConcat	0.525 ± 0.034	0.541 ± 0.053	0.526 ± 0.036	0.535 ± 0.017	0.524 ± 0.021	0.504 ± 0.051
	AutoencoderConcat	0.534 ± 0.032	0.523 ± 0.024	0.525 ± 0.016	0.518 ± 0.03	0.506 ± 0.031	**0.519 ± 0.058**
	ConsensusCosine	**0.547 ± 0.045**	0.545 ± 0.027	**0.54 ± 0.052**	**0.542 ± 0.015**	**0.545 ± 0.068**	0.483 ± 0.035
	ConsensusEuclidean	0.541 ± 0.053	**0.549 ± 0.053**	0.51 ± 0.066	0.53 ± 0.038	0.513 ± 0.073	0.475 ± 0.032

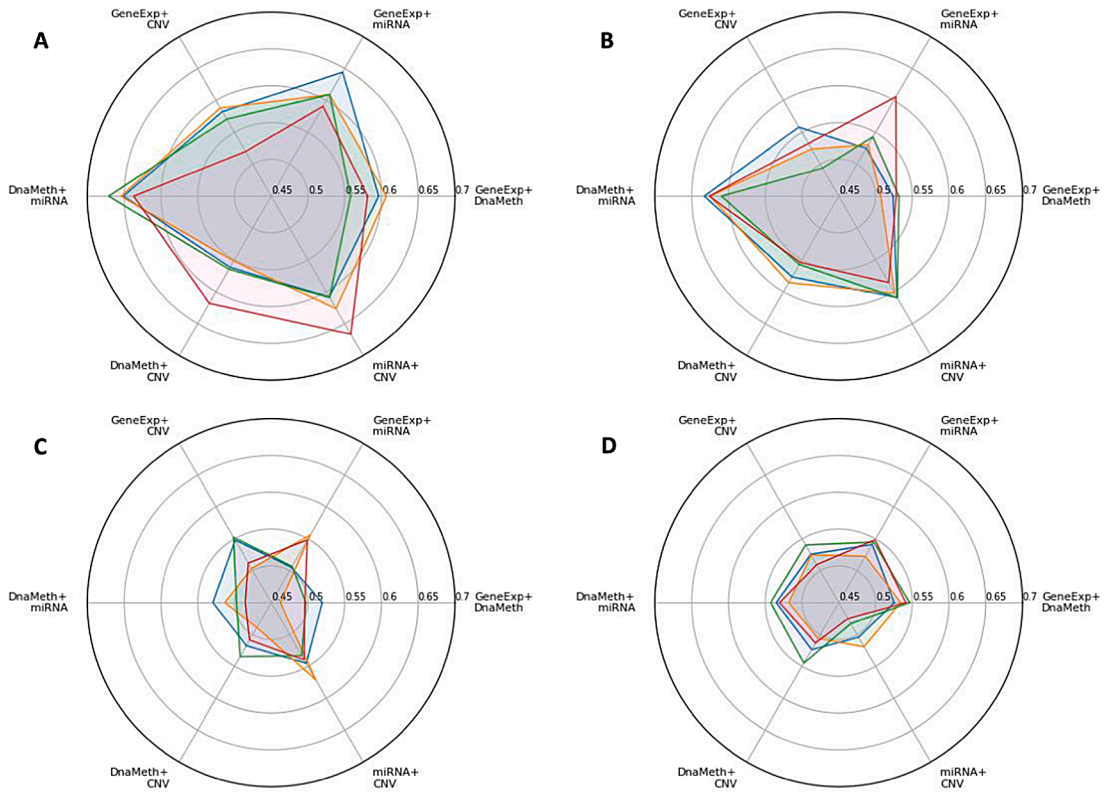

Fig. 4. Radar plot of the C-Index for overall survival analysis using concatenation and consensus based multi-omics integration. A. Breast cancer (BRCA) overall survival (OS) C-index with Top 300 PCA features. B. Breast cancer (BRCA) overall survival (OS) C-index with high variance features. C. Ovarian cancer (OV) overall survival (OS) C-index with Top 50 PCA features. D. Ovarian cancer (OV) overall survival (OS) C-index with high variance features. Green lines: SimpleConcat. Blue lines: AutoencoderConcat. Magenta lines: ConsensusCosine. Yellow lines: ConsensusEuclidean.

using the integration of the top 300 PCA features of miRNA and CNV. From this Kaplan-Meier plot, we can observe the ConsensusEuclidean achieves the best low-risk and high-risk group separation, while the SimpleConcat has the worst low-risk and high-risk groups separation. Additional Kaplan-Meier plots are presented in Supplementary Material S5.

4.2.3. *Visualization of the hidden features with t-SNE and scatter plot*

We use t-SNE to visualize the hidden features learned by four multi-omics integration methods (concatenation-based: SimpleConcat and AutoencoderConcat; consensus-based: ConsensusCosine and ConsensusEuclidean). After reducing the dimensions of the hidden features to two, we visualize the decomposed hidden features with scatter plots (Fig. 7 and Supplementary Material S6). We illustrate the integration of DnaMeth and miRNA in Fig. 7A and the integration of miRNA and CNV in Fig. 7B. From the scatter plots, we observe that the multi-omics hidden features learned by consensus-based integration align better compared to that of concatenation-based integration. The results demonstrate that the Cosine-similarity-based consensus learning (ConsensusCosine) and the Euclidean-based consensus learning (ConsensusEuclidean) improve the cross-modality alignment for multi-omics data, which contributes the performance improvements on survival analysis.

5. Discussions

In the cancer types classification, if using 1,000 high variance features, consensus-based integration consistently outperforms concatenation-based integration; while if using top 100 PCA features, concatenation-based integration consistently outperforms the consensus-based integration. Because the consensus-based integration is based on the mutual information co-existing in multiple -omics modalities, when features are more compressed as in PCA features, the concatenation strategy works better. Thus, mutual information among modalities is essential when applying consensus-based integration. For example, if integrating two distinct modalities of a patient (e.g., gene

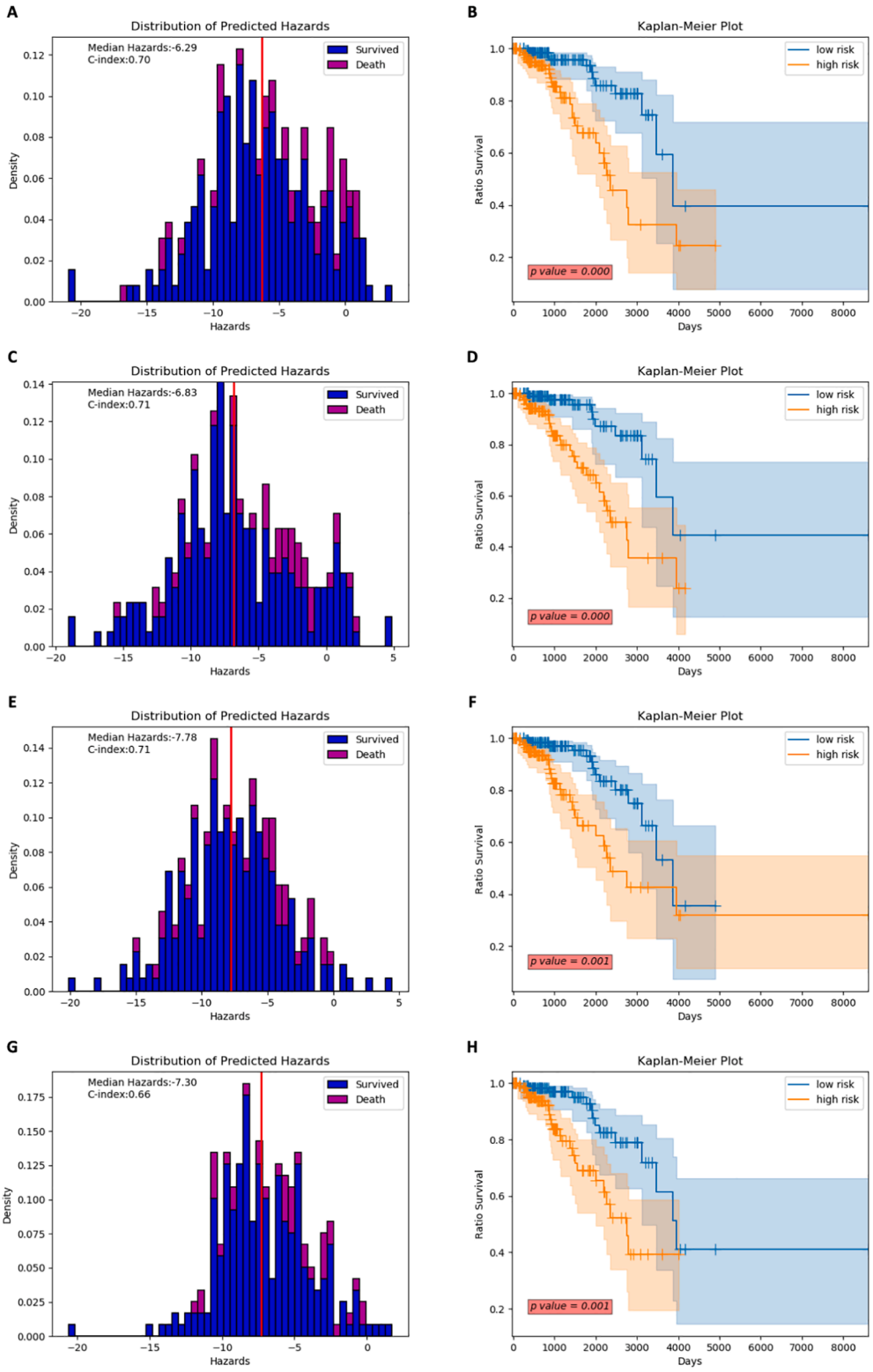

Fig. 5. Kaplan–Meier plot of the breast cancer (BRCA) overall survival (OS) prediction by integrating the top 300 PCA features of DnaMeth and miRNA on train-test split fold 1. A. Distribution of predicted hazards with SimpleConcat. B. Kaplan–Meier plot of OS prediction with SimpleConcat. C. Distribution of predicted hazards with AutoencoderConcat. D. Kaplan–Meier plot of OS prediction with AutoencoderConcat. E. Distribution of predicted hazards with ConsensusCosine. F. Kaplan–Meier plot of OS prediction with ConsensusCosine. G. Distribution of predicted hazards with ConsensusEuclidean. H. Kaplan–Meier plot of OS prediction with ConsensusEuclidean.

expression vs. MRI imaging), the consensus learning may not be a good option. In overall survival prediction, both the concatenation-based and consensus-based integration models consistently outperform the single-modality models, but the rank of their performance varies with different choices of modality combinations. A future study is needed to understand when consensus works better than concatenation in survival prediction.

In this study, we have developed a multi-omics data integration method by divergence-based modality-invariant representation. To impose consensus constraints among modalities, we maximize the Cosine similarity (ConsensusCosine) or minimize the Euclidean distances (ConsensusEucldiean) on the features represented from each -omics modality. The first future direction is to replace the Cosine similarity or Euclidean distance with other divergence criteria such as

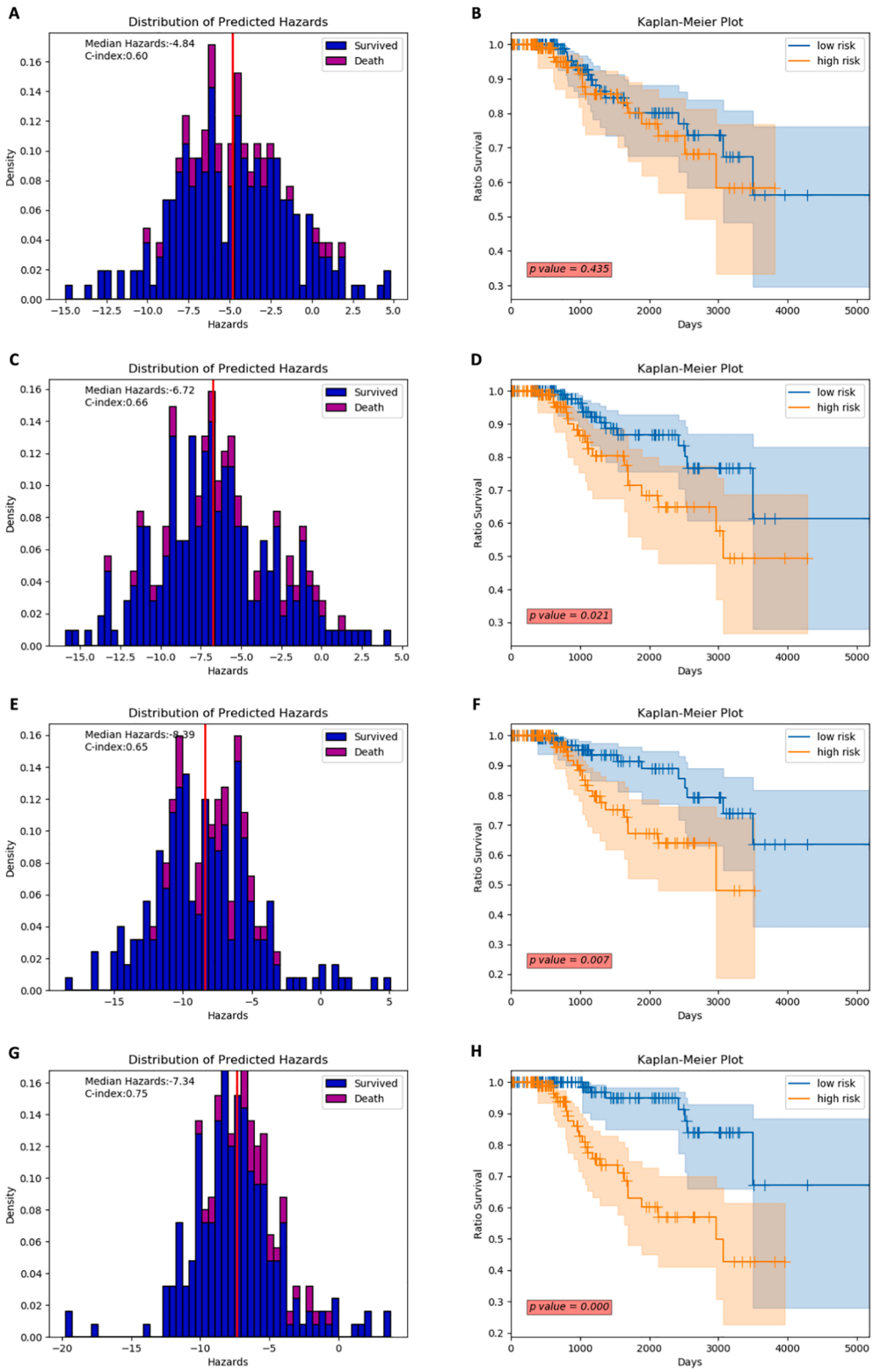

Fig. 6. Kaplan–Meier plot of the breast cancer(BRCA) overall survival (OS) prediction by integrating the top 300 PCA features of miRNA and CNV on train-test split fold 2. A. Distribution of predicted hazards with SimpleConcat. B. Kaplan–Meier plot of OS prediction with SimpleConcat. C. Distribution of predicted hazards with AutoencoderConcat. D. Kaplan–Meier plot of OS prediction with AutoencoderConcat. E. Distribution of predicted hazards with ConsensusCosine. F. Kaplan–Meier plot of OS prediction with ConsensusCosine. G. Distribution of predicted hazards with ConsensusEuclidean. H. Kaplan–Meier plot of OS prediction with ConsensusEuclidean.

Wasserstein distance, or to adopt adversarial learning for modality-invariant representation instead of the predefined divergence metrics. That is to differentiate features represented from various modalities by training the discriminator and the modality-specific encoders in an adversarial fashion.

The second future direction is to integrate the feature selection or dimension reduction step with our multi-modality network. Currently, we have utilized variance for feature selection and PCA for dimension reduction, respectively, which have shown influences on multi-modality integration performance. As deep learning can offer fully data-driven analysis, we can integrate the feature selection step with the modality-invariant representation step to improve the overall model performance. However, to enable such automatic feature engineering, one bottleneck could be the amount of data available for training.

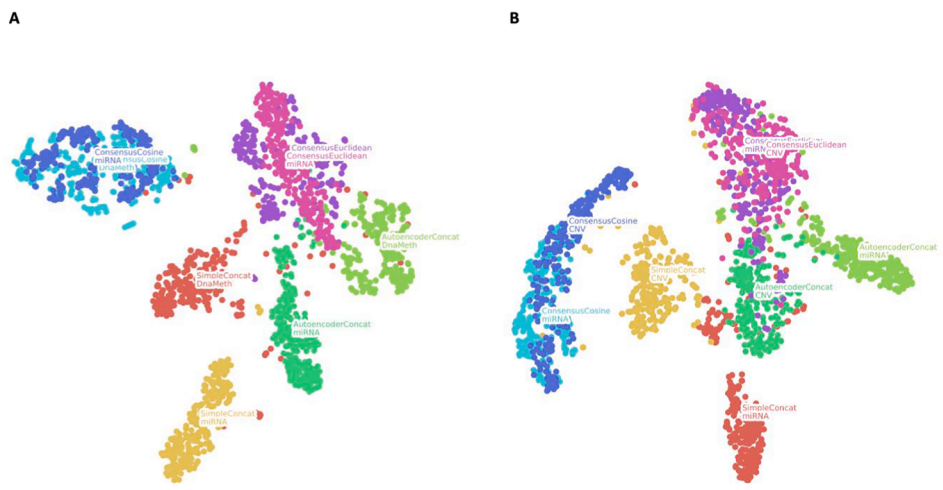

Fig. 7. Scatter Plot of the hidden features generated for breast cancer (BRCA) overall survival (OS) analysis after t-SNE decomposition. A. t-SNE scatter plot for hidden features represented from the top 300 PCA features of DnaMeth and miRNA (cross-validation fold 1). B. t-SNE scatter plot for hidden features represented from the top 300 PCA features of miRNA and CNV (cross-validation fold 2). Additional t-SNE scatter plots for the other cross-validation folds are presented in Supplementary Material S6.

Another future direction is to utilize multi-task learning with multi-view learning. In this study, we train encoders for each -omics modality (e.g., GeneExp, DnaMeth, miRNA, and CNV) independently when applying to various endpoints (e.g., cancer types classification and survival analysis). However, for encoders of a specific modality (e.g., GeneExp), they transform raw -omics features to hidden spaces with the same dimensions, despite that hidden features might be used for various cancer types (BRCA vs. OV) and prediction tasks (e.g., cancer types classification vs. survival analysis). Thus, we will apply multi-mask learning to combine the training of these encoders and obtain potential performance improvements.

6. Conclusions

In this paper, we presented an effective end-to-end solution for integrating multi-omics data based on deep neural networks. We developed a new divergence-based consensus regularization techniques to capture the consensus information among modalities to improve prediction performance. Although we only tested this new algorithm for a two-modality integration scenario with two divergence metrics (i.e., Cosine similarity and Euclidean distance), this novel method can be easily extended to multiple modalities and advanced divergence metrics. Through our analysis, we demonstrated on why and how to integrate different modalities of data. Our experiment results validated the effectiveness of our new method for both cancer types classification and survival prediction. In the visualization of features, we observed that our approach could identify more consensus among modalities with clearer separation margins.

CRediT authorship contribution statement

Li Tong: Conceptualization, Methodology, Software, Investigation, Visualization, Writing - original draft. **Hang Wu:** Investigation, Writing - original draft. **May D. Wang:** Supervision, Resources, Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The work was supported in part by grants from the National Institute of Health (NIH) under Award R01CA163256, Giglio Breast Cancer Research Fund, Petit Institute Faculty Fellow and Carol Ann and David D. Flanagan Faculty Fellow Research Fund, and Georgia Cancer Coalition Distinguished Cancer Scholar award. This work was also supported in part by the scholarship from China Scholarship Council (CSC) under the Grant CSC NO. 201406010343. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The results published here are in whole or part based upon data generated by the TCGA Research Network:https://www.cancer.gov/tcga.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.ymeth.2020.07.008.

References

[1] P. Indyk, R. Motwani, Approximate nearest neighbors: Towards removing the curse of dimensionality, in: Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, STOC '98, Association for Computing Machinery, pp. 604-613.https://doi.org/10.1145/276698.276876.
[2] Y. EL-Manzalawy, T.-Y. Hsieh, M. Shivakumar, D. Kim, V. Honavar, Min-redundancy and max-relevance multi-view feature selection for predicting ovarian cancer survival using multi-omics data 11(3) 71.https://doi.org/10.1186/s12920-018-0388-0.
[3] O.B. Poirion, K. Chaudhary, L.X. Garmire, Deep Learning data integration for better risk stratification models of bladder cancer 2018 197–206. arXiv:29888072. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5961799/.
[4] T. Ma, A. Zhang, Multi-view Factorization AutoEncoder with Network Constraints for Multi-omic Integrative Analysis, in: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 702–707.https://doi.org/10.1109/BIBM.2018.8621379.
[5] Z. Huang, X. Zhan, S. Xiang, T.S. Johnson, B. Helm, C.Y. Yu, J. Zhang, P. Salama, M. Rizkalla, Z. Han, K. Huang, SALMON: survival Analysis Learning With Multi-Omics Neural Networks on Breast Cancer 10.https://doi.org/10.3389/fgene.2019.00166. URL: https://www.frontiersin.org/articles/10.3389/fgene.2019.00166/full.
[6] J. Mitchel, K. Chatlin, L. Tong, M.D. Wang, A translational pipeline for overall survival prediction of breast cancer patients by decision-level integration of multi-omics data, in: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1573–1580.https://doi.org/10.1109/BIBM47256.2019.8983243.
[7] K. Chaudhary, O.B. Poirion, L. Lu, L.X. Garmire, Deep learning-based multi-omics integration robustly predicts survival in liver cancer 24(6) 1248–1259. arXiv:28982688,https://doi.org/10.1158/1078-0432.CCR-17-0853. URL: https://clincancerres.aacrjournals.org/content/24/6/1248.
[8] C. Xu, D. Tao, C. Xu, A survey on multi-view learning. arXiv:1304.5634.
[9] T. Baltrušaitis, C. Ahuja, L.-P. Morency, Multimodal machine learning: a survey and taxonomy. arXiv:1705.09406.
[10] M. Gönen, E. Alpaydın, Multiple kernel learning Algorithms 12, 2211–2268.
[11] Y. Song, L.-P. Morency, R. Davis, Multi-view latent variable discriminative models for action recognition, IEEE, pp. 2120–2127.
[12] K. Liu, Y. Li, N. Xu, P. Natarajan, Learn to combine modalities in multimodal deep learning. arXiv:1805.11730.

[13] W. Guo, J. Wang, S. Wang, Deep multimodal representation learning: a survey 7, 63373–63394, conference Name: IEEE Access.https://doi.org/10.1109/ACCESS.2019.2916887.
[14] J.H. Phan, C.F. Quo, C. Cheng, M.D. Wang, Multiscale integration of-omic, imaging, and clinical data in biomedical informatics, IEEE Rev. Biomed. Eng. 5 (2012) 74–87.
[15] P.-Y. Wu, R. Chandramohan, J.H. Phan, W.T. Mahle, J.W. Gaynor, K.O. Maher, M. D. Wang, Cardiovascular transcriptomics and epigenomics using next-generation sequencing: challenges, progress, and opportunities, circulation: cardiovascular, Genetics 7 (5) (2014) 701–710.
[16] C.D. Kaddi, M.D. Wang, Developing robust predictive models for head and neck cancer across microarray and rna-seq data, in: Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics, 2015, pp. 393–402.
[17] S. Mishra, C.D. Kaddi, M.D. Wang, Pan-cancer analysis for studying cancer stage using protein and gene expression data, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2016, pp. 2440–2443.
[18] Y. Li, F.-X. Wu, A. Ngom, A review on machine learning principles for multi-view biological data integration 19 (2) 325–340. doi:10.1093/bib/bbw113. URL: https://academic.oup.com/bib/article/19/2/325/2664338.
[19] M. Kim, I. Tagkopoulos, Data integration and predictive modeling methods for multi-omics datasets 14(1) 8–25.https://doi.org/10.1039/C7MO00051K. URL: https://pubs.rsc.org/en/content/articlelanding/2018/mo/c7mo00051k.
[20] A. Xu, J. Chen, H. Peng, G. Han, H. Cai, Simultaneous interrogation of cancer omics to identify subtypes with significant clinical differences. 10.https://doi.org/10.3389/fgene.2019.00236. URL: https://www.frontiersin.org/articles/10.3389/fgene.2019.00236/full.
[21] S.V. Vasaikar, P. Straub, J. Wang, B. Zhang, LinkedOmics: analyzing multi-omics data within and across 32 cancer types 46(D1) D956–D963.https://doi.org/10.1093/nar/gkx1090. URL: https://academic.oup.com/nar/article/46/D1/D956/4607804.
[22] G.P. Way, F. Sanchez-Vega, K. La, J. Armenia, W.K. Chatila, A. Luna, C. Sander, A. D. Cherniack, M. Mina, G. Ciriello, Machine learning detects pan-cancer ras pathway activation in the cancer genome atlas 23(1) 172–180.
[23] M. Goldman, B. Craft, M. Hastie, K. Repečka, F. McDade, A. Kamath, A. Banerjee, Y. Luo, D. Rogers, A.N. Brooks, J. Zhu, D. Haussler, The UCSC Xena platform for public and private cancer genomics data visualization and interpretation 326470. https://doi.org/10.1101/326470. URL: https://www.biorxiv.org/content/10.1101/326470v6.
[24] C. Meng, O.A. Zeleznik, G.G. Thallinger, B. Kuster, A.M. Gholami, A.C. Culhane, Dimension reduction techniques for the integrative analysis of multi-omics data 17 (4) 628–641.https://doi.org/10.1093/bib/bbv108. URL: https://academic.oup.com/bib/article/17/4/628/2240645.
[25] N.E. Breslow, Analysis of survival data under the proportional hazards model 43 (1) 45–57.https://doi.org/10.2307/1402659. URL: https://www.jstor.org/stable/1402659.
[26] D.R. Cox, Regression models and life-tables 34(2) 187–202, _eprint:https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.
[27] J.L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, Y. Kluger, DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network 18(1) 24.
[28] H. Kvamme, Borgan, I. Scheel, Time-to-event prediction with neural networks and Cox regression 20(129) 1–30.
[29] H. Uno, T. Cai, M.J. Pencina, R.B. D'Agostino, L. Wei, On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data 30(10) 1105–1117.
[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, Duchesnay, Scikit-learn: Machine Learning in Python 12 2825–2830. URL: http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.
[31] C. Davidson-Pilon, J. Kalderstam, P. Zivich, B. Kuhn, M. Williamson, AbdealiJK, A. Fiore-Gartland, L. Moneda, Gabriel, D. WIlson, A. Parij, K. Stark, S. Anton, M.S. Peña, L. Besson, Jona, H. Gadgil, D. Golland, S. Hussey, R. Kumar, J. Noorbakhsh, A. Klintberg, D. Medvinsky, D. Zgonjanin, D.S. Katz, D. Chen, C. Ahern, C. Fournier, A. Moncada-Torres, A.F. Rendeiro, CamDavidsonPilon/lifelines: V0.23.7. https://doi.org/10.5281/zenodo.3608331. URL: https://zenodo.org/record/3608331#.Xh-9of70nRZ.