1. *Multi-ethnic Survival Prediction: Transfer Learning with Cox Neural Network (kinda useful)*

(1) four datasets: KIRC (406EA & 44AA), GBMLGG(582EA & 48 AA), two synthetic ones

(2) Cox Neural Network (6 layers):

> Here we created a Cox neural network model with six layers. The input layer has 195 nodes for KIRC and 176 nodes for GBMLGG, a fully connected (FC) layer with 128 nodes followed by a dropout layer (drop out rate = 0.5), then another FC layer (with 64 nodes) also followed by a dropout layer (p = 0.5), and finally a Cox regression layer. We used the ReLU activation function for each FC layer to avoid the gradient vanish problem Goodfellow et al. (2016). In model fitting, we optimized the object function, $l(\beta) = -\sum_{i \epsilon U} L(\beta) + \lambda_1 |W| + \lambda_2 ||W||_2$, where $L(\beta) = log[\prod_{i=1}^{m} \frac{e^{\beta X_i}}{\sum_{j \epsilon R(T_i)} e^{\beta X_j}}]^{\delta_i} = \sum_{i=1}^{m} \delta_i [\beta X_i - log\left\{\sum_{j \epsilon R(T_i)} e^{\beta X_j}\right\}]$, β represents the weights of the Cox layer, $\sum_{i \epsilon U} L(\beta)$ is the partial likelihood, U is the set of uncensored patients, λ_1 and λ_2 are regularization parameters, W represents the weights in the network, δ_i is the event status of patient i.

(3) Transfer Learning Strategies:

First one: pre-trained with EA data and fine-tuned with AA

Second one:

> learning rate (0.002) since the model had been partially fitted. (2) The second fine-tuning method is based on stacked auto-encoder Sevakula et al. (2018); Singh et al. (2016); Vincent et al. (2010). We used the unlabeled data from the EA group to pre-train a stacked denoising auto-encoder with 5 layers: an input layer, a FC layer with 128 nodes, a bottleneck layer with 64 nodes, a FC layer with 128 nodes and an output layer with same nodes as the input layer.

(4) Evaluation Method: C-index

(5) Result:

Table 1: Multi-ethnic Machine Learning Scheme Comparison (*SD: Synthetic data)

Multi-ethnic ML Scheme	Experiment	Ethnic Composition		C-index			
		Training Data	Testing Data	KIRC	GBMLGG	SD1*	SD2*
Mixture Learning	Mixture 0	AA + EA	AA + EA	0.68	0.74	0.81	0.86
	Mixture 1		EA	0.69	0.75	0.83	0.87
	Mixture 2		AA	0.52	0.59	0.56	0.65
Independent Learning	Independent 1	EA	EA	0.69	0.75	0.85	0.89
	Independent 2	AA	AA	0.43	0.64	0.55	0.54
Naive Transfer	Naive Transfer	EA	AA	0.45	0.63	0.51	0.64
Transfer Learning	Transfer Learning	EA (source) AA (target)	AA	0.66	0.69	0.73	0.69

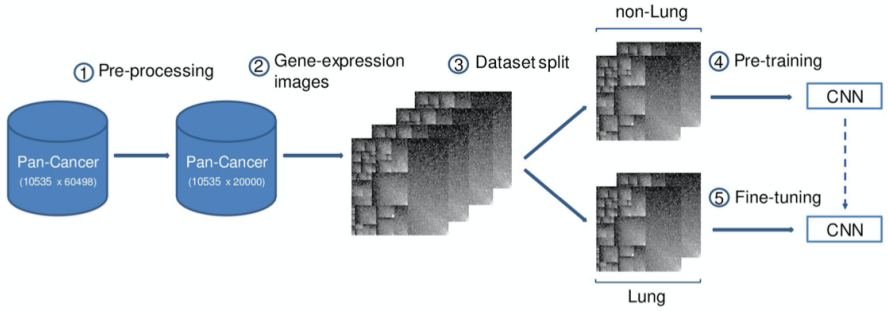
2. *Transfer learning with convolutional neural networks for cancer survival prediction using gene-expression data (quite a different way)*

(1) Procedure:

*pre-processing: feature reduction

*gene-expression images: use KEGG BRITE functional hierarchies to map from KEGG BRITE database information (existed) to the genes contained in the Pan-Cancer dataset; construct the images based on the genes achieved from the last sub step and recursively divided the images

*transfer learning: non-lung dataset (combination of the rest sub datasets except lung datasets) for pre-training, lung dataset (combination of lung datasets) for fine-tuning



(2) Nota Bene:
*all the right-censored samples whose censoring times are below time t must be discarded
*in their experiment, they chose 230 days for the time t to get trade-off between imbalanced datasets and the number of samples
*to alleviate the effects of imbalance, random over-sampling used
*AUC as evaluation metric

3. *Deep Transfer Learning and Radiomics Feature Prediction of Survival of patients with High-Grade Gliomas (useless)*

(1) About survival prediction based on the medical images (including segmentation…)
(2) So has nothing to do with our topic

4. *Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma (useless)*

(1) Same as the third paper, it's more about image classification

5. *TLSurv: Integrating Multi-Omics Data by Multi-stage Transfer Learning for Cancer Survival Prediction (useful)*

(1) Dataset: LUAD dataset with five data modalities (copy number variation/DNA methylation/miRNA expression/mRNA expression and survival outcomes)
(2) Procedure:

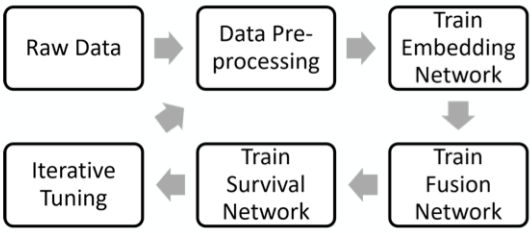


Figure 1: Overall Pipeline for Development of TLSurv. Raw data were first preprocessed for quality control. The cleaned data were fed into training the embedding networks of each modality for dimensionality reduction. The fusion network was then trained, followed by the survival network for predictions. The entire training pipeline was iterated to optimize the performance.

(3) Network architecture:

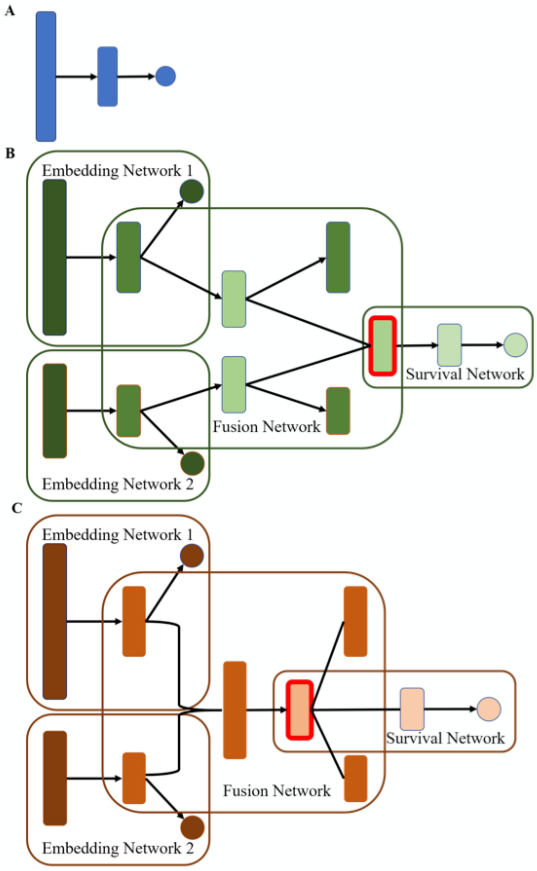


Figure 3: Architectures of Cox-nnet, TLSurv(MAE) and TLSurv(VAE). A. Cox-nnet with one hidden layer; B. TLSurv with MAE implementation for two modalities; C. TLSurv with VAE implementation for two modalities; For B and C, the four bounding boxes represent the four individual networks within the super-hybrid network. Leftmost rectangular layers correspond to inputs from different data sources. Rectangles with thick red outlines represent fused views. All seven circular output nodes represent log-partial hazard, while the four rectangular output nodes represent reconstructed embedding representations. Arrows represent dense connections between adjacent layers, and the right brace indicates concatenation followed by dense connections. The two lines in B refers to affinity network fusion, while the three lines in C refers to sampling from the learned distribution.

Figure A Cox-nnet as baseline model. In Figure B and C, then first embedding section will reduce dimensionality. Fusion Network will merge the data modalities while the last survival network's output neuron is a classic log-partial hazard function.

(4) Evaluation metric: C-index

(5) Summary of novelty:

2.9 Summary of novelty

Overall, our contributions can be summarized as follows. To our knowledge, this is the first deep-learning based integrative survival model for lung cancer.

In order to address the "curse of dimensionality" issue, we built TLSurv on a super-hybrid network architecture with four individual neural networks. Using a super-hybrid network to integrate multiple -omics data is new to the field, and TLSurv can be applied to other heterogeneous data modalities. As a modular network, many different embedding networks for different types of input data may be used. For example, a convolutional neural network can be trained to integrate clinical images alongside -omic data, or a recurrent neural network can be trained to integrate sequential data.

We also developed a novel multi-stage transfer learning scheme to train the super-hybrid network incrementally. Each training stage has its own purpose, and it inherits the weights from previously trained sections with the exception of the first stage. As a result, we are effectively training a shallow rather than deep neural network during each training stage, but a powerful deep neural network will be obtained at the end. Since only a shallow neural network is trained during each stage, it mitigates the issues of vanishing gradients[18] and exploding gradients[34].

Lastly, rather than using traditional Gene Set Enrichment Analysis (GSEA), we have used the primary attribution algorithm for interpretability analysis to identify those significant input features. In this way, novel biomarkers could be identified instead of verifying those well established pathways. Those findings can potentially provide new insights for medical and pharmaceutical professionals.

6.