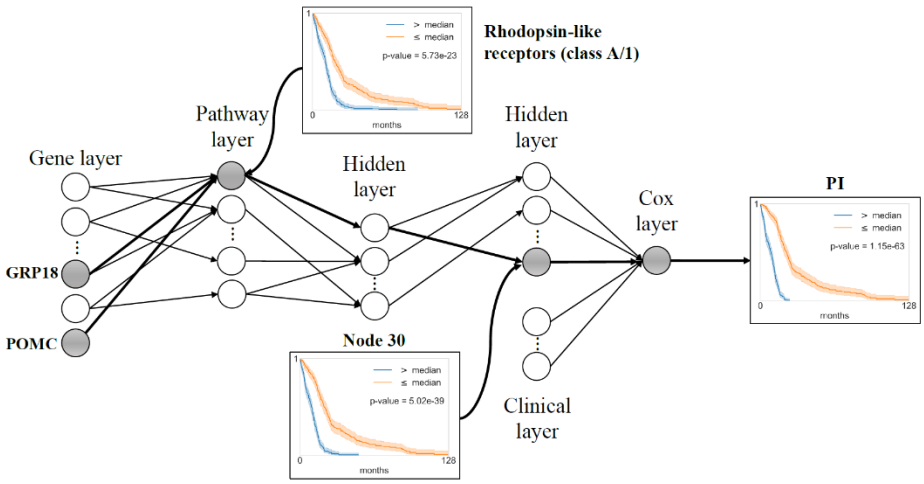
Data Science Research Series

Cox-PASNet: Pathway-based Sparse Deep Neural Network for Survival Analysis

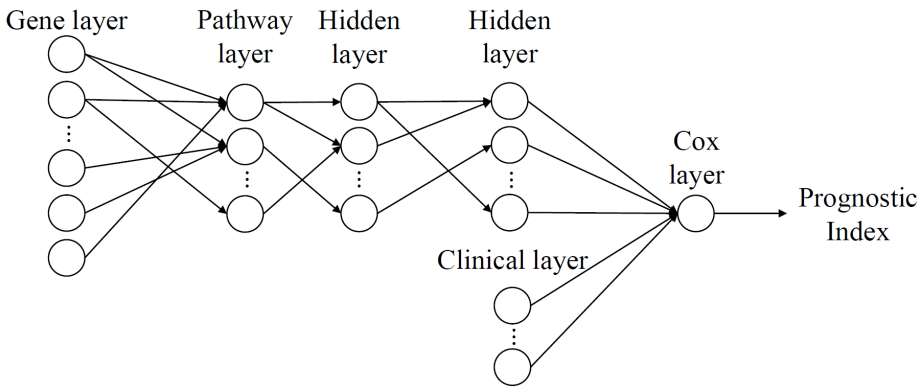


BACKGROUND

An in-depth understanding of complex biological processes associated to patients' survival time at the cellular and molecular level is critical not only for developing new treatments for patients but also for accurate survival prediction. However, highly nonlinear and high-dimension, low-sample size (HDLSS) data cause computational challenges in survival analysis. We developed a novel pathway-based, sparse deep neural network, called Cox-PASNet, for survival analysis by integrating high-dimensional gene expression data and clinical data. Cox-PASNet is a biologically interpretable neural network model where nodes in the network correspond to specific genes and pathways, while capturing nonlinear and hierarchical effects of biological pathways to a patient's survival. We also provide a solution to train the deep neural network model with HDLSS data.

APPROACH

We introduce our proposed model, Cox-PASNet, a pathway-based sparse deep neural network for survival analysis with genomic and clinical data. Cox-PASNet combines a Cox proportional hazards regression with a deep neural network, incorporating prior knowledge of biological pathways. The architecture of Cox-PASNet is comprised of (1) a gene layer, (2) a pathway layer, (3) multiple hidden layers, (4) a clinical layer, and (5) a Cox layer.



The gene layer is an input layer of Cox-PASNet introducing gene expression data with n patient samples of p gene expressions. The pathway layer represents biological pathways where a node indicates a specific biological pathway. The hidden layers model the nonlinear and hierarchical effects of pathways. The clinical layer introduces clinical data to the model separately from genomic data. The Cox layer is the output layer that has only one node, which is introduced to a Cox-PH model.

In order to perform Cox proportional hazards regression on the Cox layer, Cox-PASNet defines the objective function using average negative log partial likelihood with L^2 regularization. The model is optimized by partially training small sub-networks with sparse coding. The algorithm of Cox-PASNet is briefly described in the algorithm below:

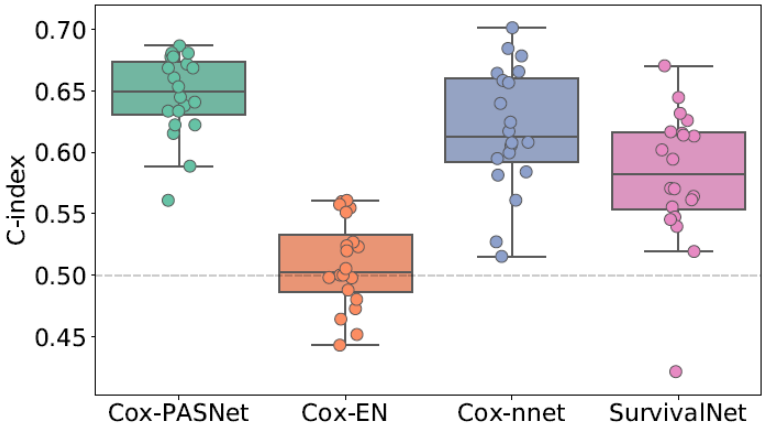
Algorithm 1 Training of Cox-PASNet

1: Initialize weights $\mathbf{W}^{(\ell)}$, biases $\mathbf{b}^{(\ell)}$, and $\boldsymbol{\beta}$
2: $\mathbf{W}^{(0)} \leftarrow \mathbf{W}^{(0)} \star \mathbf{M}^{(0)}$
3: **repeat**
4: Select a small sub-network via dropout
5: Train the sub-network
6: Sparse coding with the optimal $\mathbf{M}^{(\ell)}$ by Eq. (3)
7: Update weights
8: **until** convergence

RESULTS

In this study, we considered glioblastoma multiforme (GBM) cancer to assess Cox-PASNet. Gene expression and clinical data of GBM were obtained from the Cancer Genome Altas (TCGA, http://cancergenome.nih.gov). The prior knowledge of biological pathways were taken from the Molecular Signatures Database (MSigDB). We used 5,567 genes, 860 pathways, and clinical data of age from 522 GBM samples.

Cox-PASNet was evaluated by comparing the performance of different cutting-edge survival methods such as Cox-nnet, SurvivalNet, and Cox elastic net (Cox-EN). The performance of the four models were evaluated by C-index. The experimental results are illustrated below:



Model	C-index
Cox-PASNet	**0.6463 ± 0.0332**
Cox-EN	0.5089 ± 0.0362
Cox-nnet	0.6187 ± 0.0507
SurvivalNet	0.5811 ± 0.0546

Cox-PASNet showed the highest C-index of 0.6463. The statistical significance of the performance was assessed by Wilcoxon rank-sum test. The distributions of c-index scores produced by Cox-PASNet were significantly higher than others in the table below:

	Wilcoxon rank-sum test
Cox-PASNet vs. Cox-EN	8.8574e-05*
Cox-PASNet vs. Cox-nnet	0.0365*
Cox-PASNet vs. SurvivalNet	0.0008*

* shows the statistical significance with significance level = 0.05.

CONCLUSIONS

We developed a pathway-based sparse deep neural network, Cox-PASNet, for survival analysis coupled with Cox-PH model on a deep neural network. Cox-PASNet builds a neural network model that can describe nonlinear and hierarchical effects of biological pathways and provides significant prognostic factors for accurate prediction of patients' survival. A new strategy to train the deep neural network model with HDLSS data is also introduced in the paper. Cox-PASNet outperformed the current cutting-edge survival methods such as Cox-nnet, SurvivalNet, and Cox-EN, and its predictive performance was statistically assessed.

CITATION FOR FULL ARTICLE

Hao J, Kim Y, Mallavarapu T, Oh JH, Kang M

Cox-PASNet: Pathway-based Sparse Deep Neural Network for Survival Analysis

Accepted as a regular paper in IEEE International Conference on Bioinformatics & Biomedicine (IEEE BIBM 2018)

Acceptance rate: 19.6%, 105 out of 534