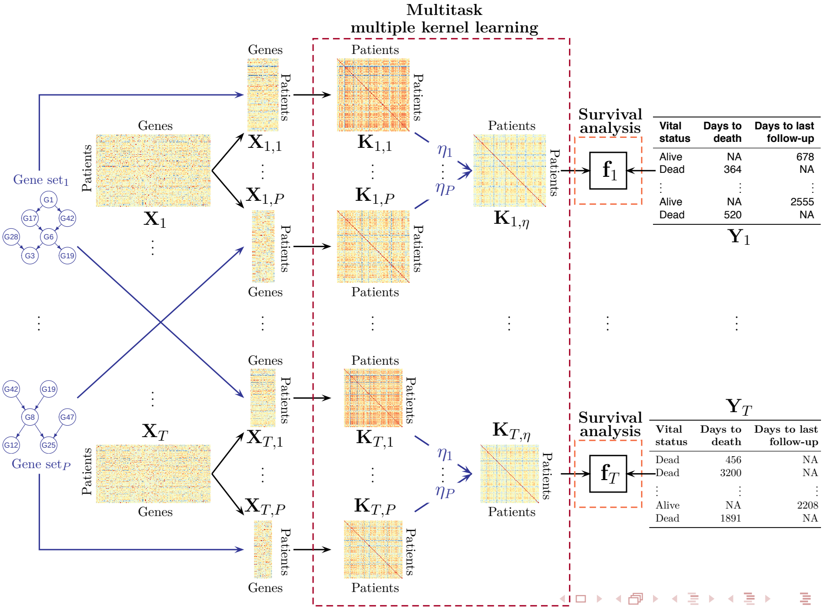
A Multitask Multiple Kernel Learning Algorithm for Survival Analysis with Application to Cancer Biology

Onur Dereli, Ceyda Oğuz, Mehmet Gönen

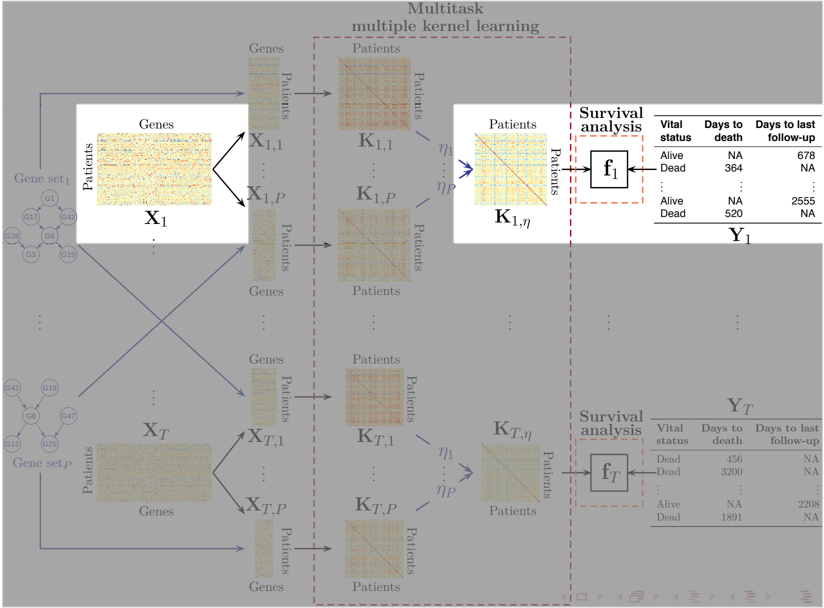
Koç University, İstanbul, Turkey
odereli14@ku.edu.tr coguz@ku.edu.tr mehmetgonen@ku.edu.tr

June 12, 2019 / Long Beach

Proposed Approach



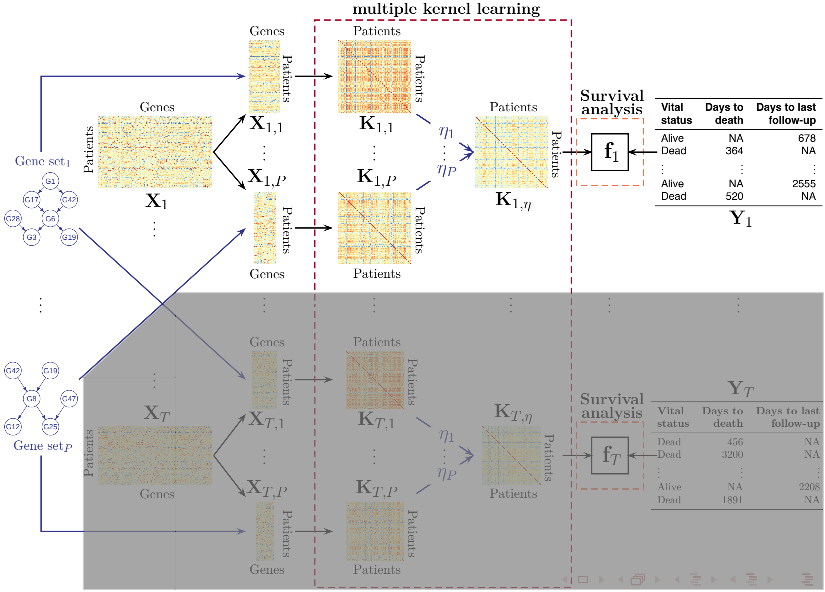
Proposed Approach



Vital status	Days to death	Days to last follow-up
Alive	NA	678
Dead	364	NA
⋮	⋮	⋮
Alive	NA	2555
Dead	520	NA

Vital status	Days to death	Days to last follow-up
Dead	456	NA
Dead	3200	NA
⋮	⋮	⋮
Alive	NA	2208
Dead	1891	NA

Proposed Approach



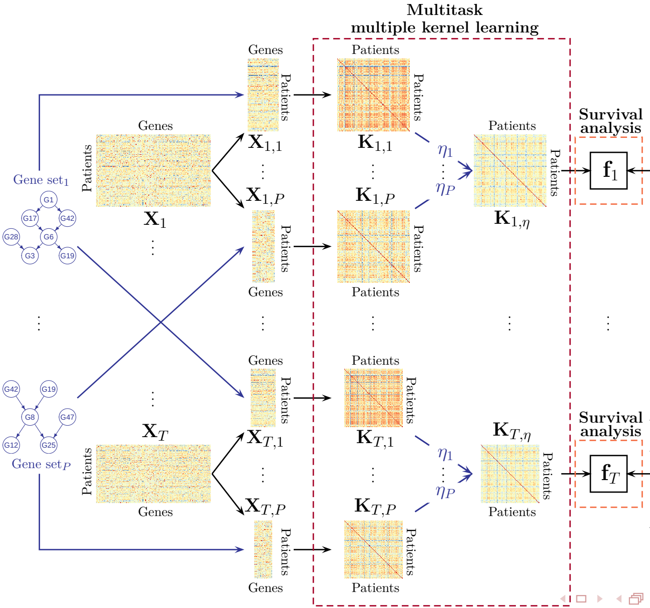
Vital status	Days to death	Days to last follow-up
Alive	NA	678
Dead	364	NA
⋮	⋮	⋮
Alive	NA	2555
Dead	520	NA

$\mathbf{Y}_1$

$\mathbf{Y}_T$

Vital status	Days to death	Days to last follow-up
Dead	456	NA
Dead	3200	NA
⋮	⋮	⋮
Alive	NA	2208
Dead	1891	NA

Proposed Approach



Vital status	Days to death	Days to last follow-up
Alive	NA	678
Dead	364	NA
⋮	⋮	⋮
Alive	NA	2555
Dead	520	NA

$\mathbf{Y}_1$

$\mathbf{Y}_T$

Vital status	Days to death	Days to last follow-up
Dead	456	NA
Dead	3200	NA
⋮	⋮	⋮
Alive	NA	2208
Dead	1891	NA

Multitask Survival MKL Formulation

$$
\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} \left[\frac{1}{2} \boldsymbol{w}_t^\top \boldsymbol{w}_t + C \sum_{i=1}^{N_t} (\xi_{ti}^{+} + (1 - \delta_{ti}) \xi_{ti}^{-}) \right] \\
\text{with respect to} \quad & \boldsymbol{w}_t \in \mathbb{R}^{D_t}, \quad \boldsymbol{\xi}_t^{+} \in \mathbb{R}^{N_t}, \quad \boldsymbol{\xi}_t^{-} \in \mathbb{R}^{N_t}, \quad b_t \in \mathbb{R} \\
\text{subject to} \quad & \epsilon + \xi_{ti}^{+} \geq y_{ti} - \boldsymbol{w}_t^\top \boldsymbol{x}_{ti} - b_t \quad \forall (t, i) \\
& \epsilon + \xi_{ti}^{-} \geq \boldsymbol{w}_t^\top \boldsymbol{x}_{ti} + b_t - y_{ti} \quad \forall (t, i) \\
& \xi_{ti}^{+} \geq 0 \quad \forall (t, i) \\
& \xi_{ti}^{-} \geq 0 \quad \forall (t, i)
\end{aligned}
$$

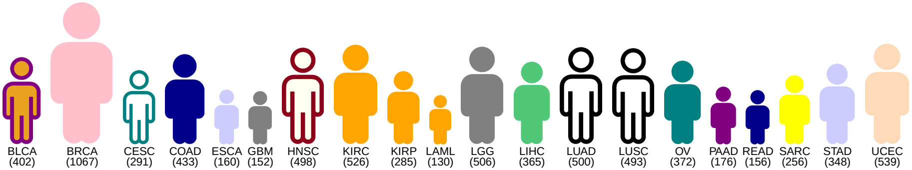
Multitask Survival MKL Formulation

$$\begin{aligned}
\text{minimize} \quad & -\sum_{t=1}^{T}\sum_{i=1}^{N_t} y_{ti}(\alpha_{ti}^{+} - \alpha_{ti}^{-}) + \epsilon \sum_{t=1}^{T}\sum_{i=1}^{N_t}(\alpha_{ti}^{+} + \alpha_{ti}^{-}) \\
& + \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\sum_{j=1}^{N_t}(\alpha_{ti}^{+} - \alpha_{ti}^{-})(\alpha_{tj}^{+} - \alpha_{tj}^{-})\sum_{m=1}^{P}\eta_m k_m(\boldsymbol{x}_{ti}, \boldsymbol{x}_{tj}) \\
\text{with respect to } & \boldsymbol{\alpha}_t^{+} \in \mathbb{R}^{N_t}, \;\; \boldsymbol{\alpha}_t^{-} \in \mathbb{R}^{N_t}, \;\; \boldsymbol{\eta} \in \mathbb{R}^{P} \\
\text{subject to} \quad & \sum_{i=1}^{N_t}(\alpha_{ti}^{+} - \alpha_{ti}^{-}) = 0 \quad \forall t \\
& C \geq \alpha_{ti}^{+} \geq 0 \quad \forall (t, i) \\
& C(1 - \delta_{ti}) \geq \alpha_{ti}^{-} \geq 0 \quad \forall (t, i) \\
& \sum_{m=1}^{P} \eta_m = 1 \\
& \eta_m \geq 0 \quad \forall m
\end{aligned}$$

Multitask Survival MKL Formulation

$$\eta_m^{(s+1)} = \frac{\sum_{t=1}^{T} \eta_m^{(s)} \sqrt{\sum_{i=1}^{N_t} \sum_{j=1}^{N_t} \alpha_{ti}^{(s)} \alpha_{tj}^{(s)} k_m(\boldsymbol{x}_{ti}, \boldsymbol{x}_{tj})}}{\sum_{t=1}^{T} \sum_{o=1}^{P} \eta_o^{(s)} \sqrt{\sum_{i=1}^{N_t} \sum_{j=1}^{N_t} \alpha_{ti}^{(s)} \alpha_{tj}^{(s)} k_o(\boldsymbol{x}_{ti}, \boldsymbol{x}_{tj})}} \quad \forall m$$

Data Sets

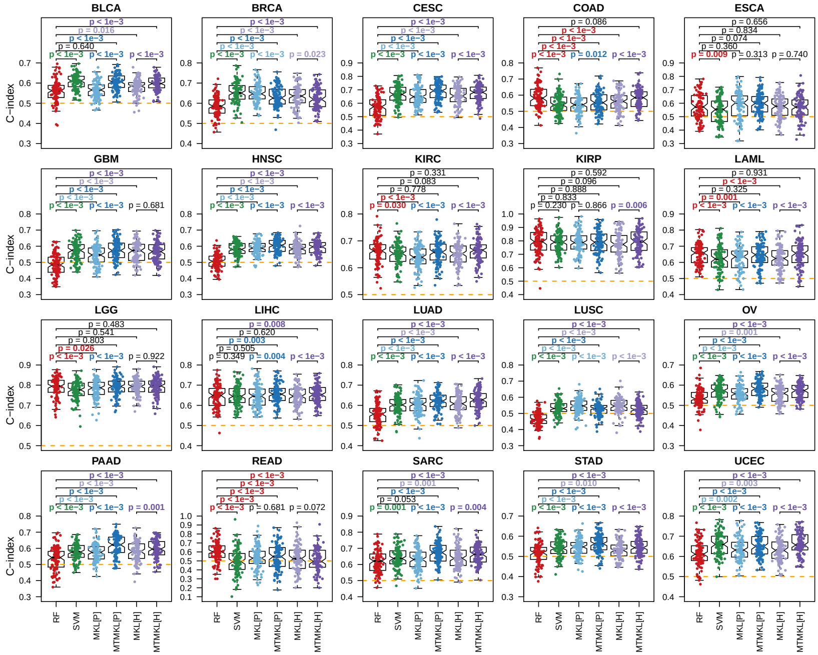


- 20 cancer data sets from TCGA database

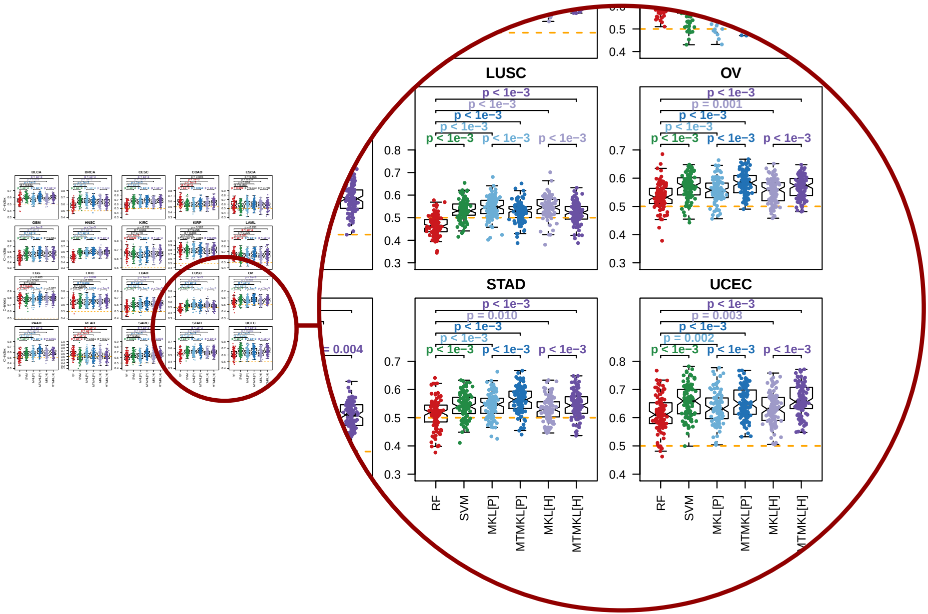
  Gene expression profiles and **survival characteristics**

- Hallmark Gene Set [1] & PID Pathway [2] Collections

KOÇ UNIVERSITY

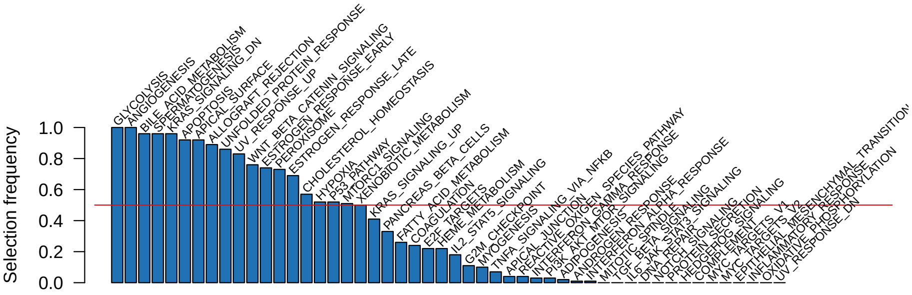


KOÇ UNIVERSITY



KOÇ UNIVERSITY

Hallmark Gene Sets



Summary

- A multitask multiple kernel learning algorithm
- Integration of different data sets
- Performing survival prediction and knowledge extraction conjointly
- Understanding the underlying mechanisms of cancer

References

[1] A. Liberzon, C. Birger, H. Thorvaldsdottir, M. Ghandi, J. P. Mesirov, and P. Tamayo. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*, 1:417–425, 2015.

[2] C. F. Schaefer, K. Anthony, S. Krupa, J. Buchoff, M. Day, et al. PID: The Pathway Interaction Database. *Nucleic Acids Res.*, 37:D674–D679, 2009.

Thank you

You can reach R implementations of our work at
`https://github.com/mehmetgonen/path2msurv`.

Please visit our poster if you have any questions.

`Room: Pacific Ballroom #242`

`Date: Wed Jun 12th 06:30 - 09:00 PM`

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant EEEAG 117E181 and Ph.D. scholarship (2211).