Personalized Donor-Recipient Matching for Organ Transplantation

JINSUNG YOON†, AHMED M. ALAA†, MARTIN CADEIRAS†, & MIHAELA VAN DER SCHAAR†,*

†UCLA, * UNIVERSITY OF OXFORD

OBJECTIVES

Organ transplants

- Therapy of choice for patients with end stage diseases!
- **National Kidney Foundation: 121,678** people waiting for lifesaving organ transplants in the U.S. as of **1/11/2016**!

Challenges

- **Post-operative complications:** infection, rejection and malignancy (Huynh 2014).
- Complications are highly dependent on the features of both **recipients** and **donors**!
- Proper recipient-donor matching requires accurate pre-operative survival analysis → **very little domain knowledge!**

Our goal → learn recipient-donor compatibility from the electronic health record data (EHR)!

STATE-OF-THE-ART

Current clinical practice

- **Donor Risk Index (DRI)** (Feng 2006)
- **Risk-Stratification Score (RSS)** (Sorror 2007)
- **Index for Mortality Prediction After Cardiac Transplantation (IMPACT)** (Weiss 2011)

Drawback of clinical risk scores:

- Expert-based, no rigorous validation.
- Ignores the heterogeneity of the recipient-donor characteristics.

Need a *data-driven* predictive model that discovers subgroups of "similar patients!

Related Machine Learning Algorithms

- ☐ **Ensemble Methods** (Kuznetsov 2014)
- ☐ **Clustering Methods** (Sontag 2016)
- ☐ **Decision Tree Methods** (Strobl 2009)

OUR METHOD

- **Main idea:** cluster the feature space into subgroups and assign a separate predictive model to every subgroup **[Build a tree with a predictor assigned to every leaf!]**.
- **Model:** • Recipient-donor feature space $\mathcal{X}$ and label space $\mathcal{Y}$. • K feature clusters $\{\mathcal{X}_1, \ldots, \mathcal{X}_K\}$.
- Cluster-specific predictors $\{h_1, \ldots, h_K\}$.
- **Learning:** Jointly optimize K, $\{\mathcal{X}_1, \ldots, \mathcal{X}_K\}$ and $\{h_1, \ldots, h_K\}$ **[Consider the Bias-Variance trade-off!]**

$$\min_{\{\mathcal{X}_1,\ldots,\mathcal{X}_k\}} \left[\min_{h_1,\ldots,h_k \in \mathcal{H}} \sum_{i=1}^{k} \mathcal{F}(\boldsymbol{X} \in \mathcal{X}_i) \times \mathbb{E}_{\mathcal{F}_i}[l(h_i(\mathbf{x}), y)] \right]$$

$$\text{subject to} \quad \mathcal{X} = \bigcup_{i=1}^{k} \mathcal{X}_i, \text{ and } \mathcal{X}_i \cap \mathcal{X}_j = \emptyset \; \forall i \neq j.$$

- **NP hard! Need an approximate solution.**

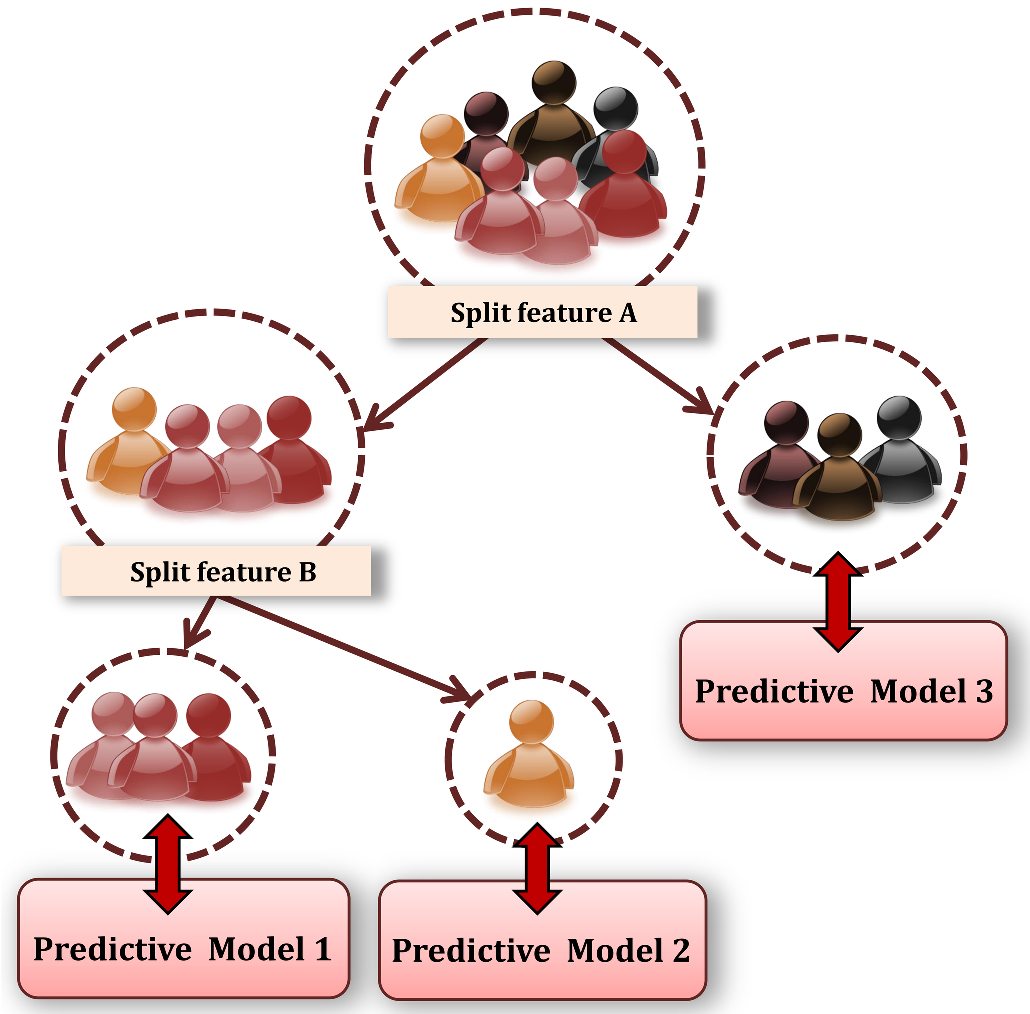
  - Use an **upper bound** on the expected error as the objective function.
  - Add **2 constraints** to make the problem tractable: Restrict clusters to hyper-cubes, & Restrict the number of clusters.

$$\min_{\{\mathcal{X}_1,\ldots,\mathcal{X}_k\}} \quad \mathcal{E} + \underbrace{\alpha\sqrt{\frac{k^2 \log M}{n}}}_{\text{Penalty term}}, \quad \mathcal{E}: \text{Empirical error}$$

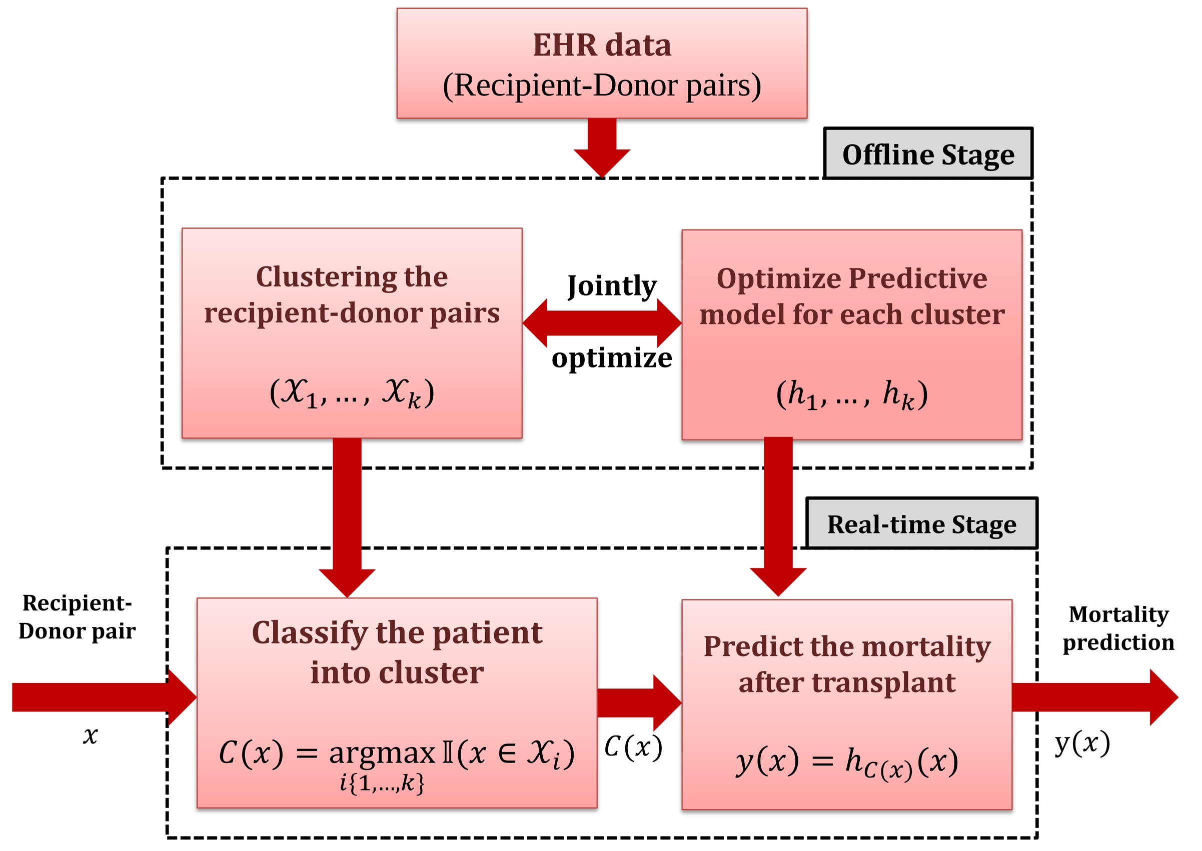
$$\text{subject to} \quad \mathcal{X}_i = \prod_{j=1}^{D} [a_{ij}, b_{ij}], a_{ij} \leq b_{ij}, a_{ij} \in \mathbb{R}^*, b_{ij} \in \mathbb{R}^*$$

$$k \leq \gamma, \text{ where } k \in \mathbb{Z}_+$$

$$\mathcal{X} = \bigcup_{i=1}^{k} \mathcal{X}_i, \text{ and } \mathcal{X}_i \cap \mathcal{X}_j = \emptyset \text{ for } \forall i \neq j.$$



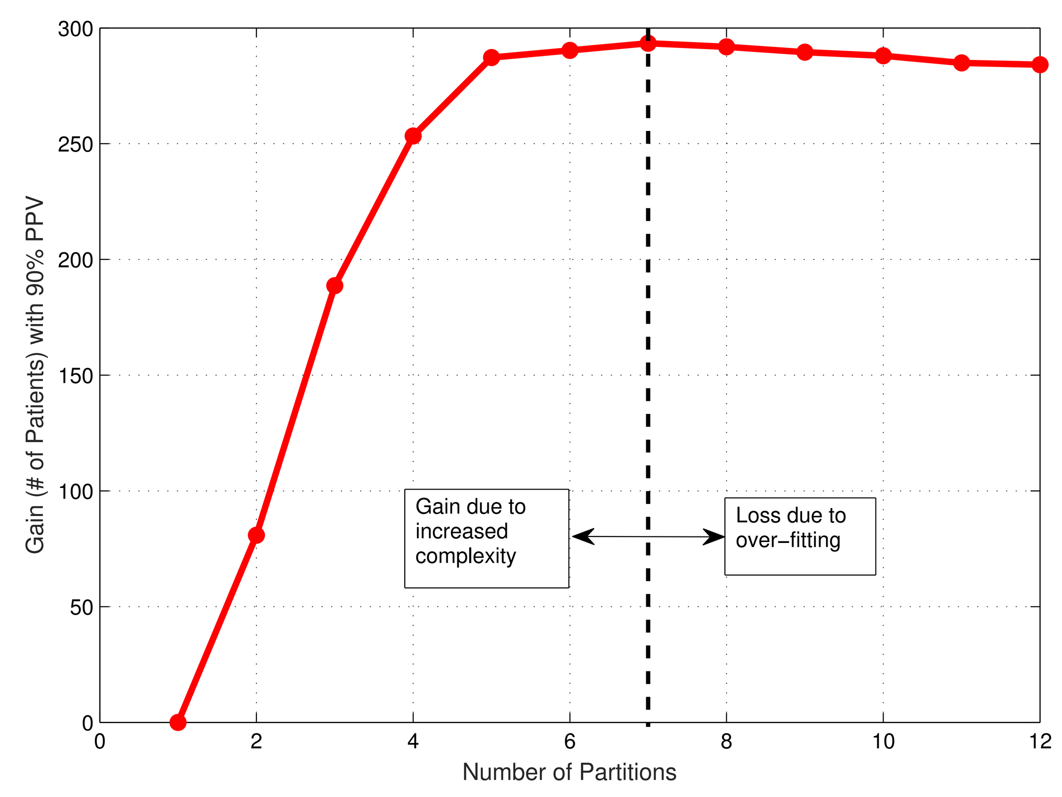
ALGORITHM

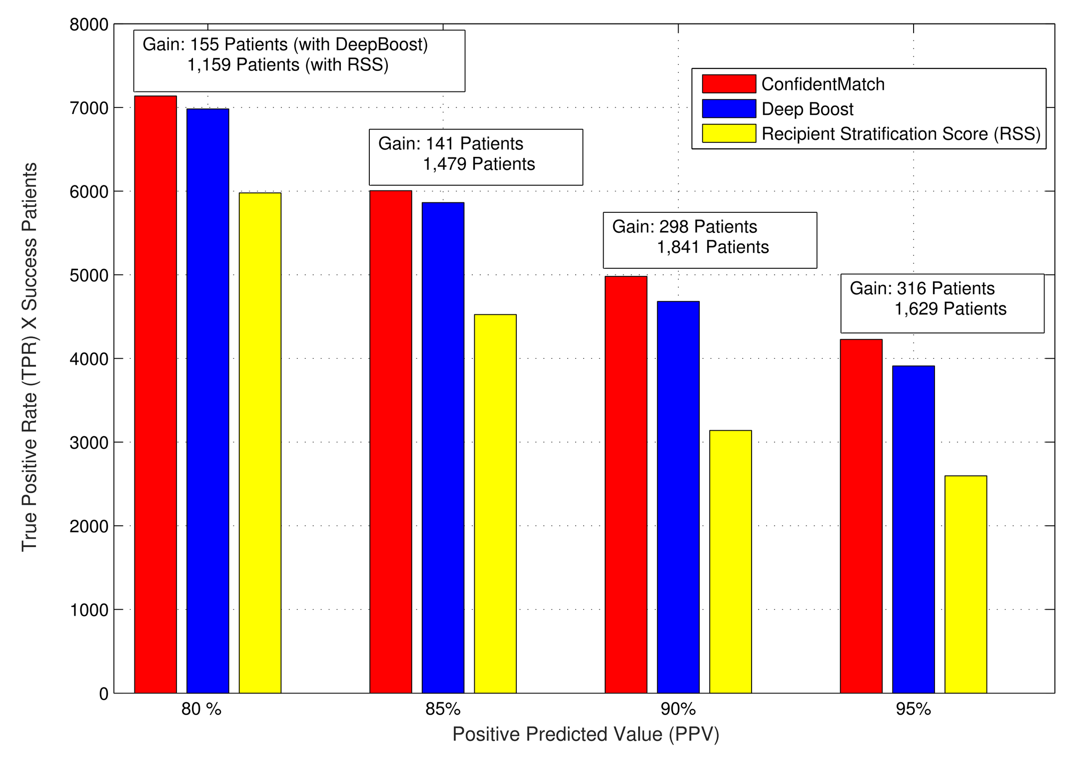


EXPERIMENTS

- United Network for Organ Sharing **(UNOS)** dataset: a cohort of **56,716** patients who got heart transplants from 1985 to 2015.
- **Training set:** transplants before **2010**.
- **Testing set:** transplants after **2010**.
- Our algorithm discovers **7 clusters (phenotypes) for the heart transplant recipient-donor pairs**.
- Gains in the number of patients for whom the PPV is 90% PPV are:
  - **298** compared to DeepBoost.
  - **1,841** compared to RSS.

RESULTS





CONCLUSIONS

- The outcomes of organ-transplant surgeries depend crucially on the individual traits of recipients and donors.
- We developed a *personalized* prognostic tools that learns a tree of predictors, the output of which is a set of recipient-donor feature clusters (phenotypes), and a predictive model customized for every cluster. The performance of our algorithm outperforms state-of-the-art clinical risk scores and other machine learning algorithms.