



Overall Survival Prediction of Breast Cancer Patients

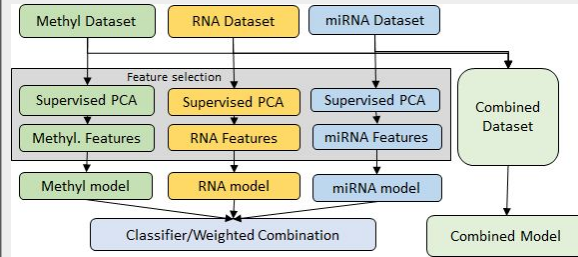
Kevin Erazo (@kerazo), Sameer Merchant (@smerchan) and Rudra S. Bandhu (@rsbandhu)



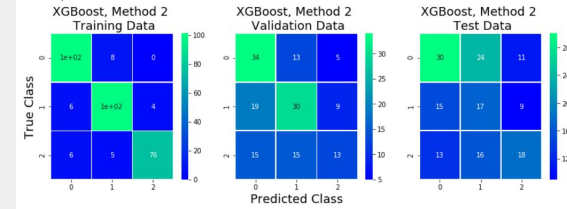
Problem

Cancer is a notoriously heterogeneous disease. The standard of care could be greatly improved with more specific molecular profiling of patients. Accurate diagnosis and tracking of cancer patients depends on using **both** clinical and molecular data to more precisely tune treatments. To that end, we developed an overall survival classifier for patients with invasive breast carcinoma.

Model Architecture & Results



Model-Based Integration using data-specific models results in a more accurate ensemble classifier^[2].



Hyperparameters	
Max Depth	2
Learning Rate	0.06-0.18
Gamma Parameter	6-9
Sub-Sampling by Rows	0.8
Sub-Sampling by Columns	0.54-0.8
Min. Child Weight	5-9

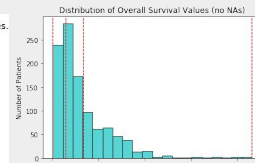
Model	Train	Valid	Test
Combined Dataset XGBoost Decision Tree	90.5%	50.3%	42.5%
Supervised PCA Stacked Ensemble	97.13%	93.37%	41.83%

Datasets

All datasets came from the TCGA-BRCA cohort stored in the LinkedOmics database^[1].

Table 1. Splitting available data into training/validation/test datasets after dropping missing values.

	Training	Validation	Test	Total Patients	Features
Clinical	749	153	153	1055	19
Methylation	461	153	153	767	17909
RNASeq	745	153	153	1051	20156
miRNA	436	153	153	742	824



Future Work

- Use class-balanced loss function to improve generalization on long term survival predictions^[3].
- Incorporate clinically-relevant metrics like IPW into models^[4].
- Explore neural networks as an alternative to an ensemble model^[5].

References

- [1] Vasaikar, S., Straub, P., Wang, J. and Zhang, B. (2017). LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Research*, 46(D1), pp.D956-D963.
- [2] Lin, E. and Lane, H. (2017). Machine learning and systems genomics approaches for multi-omics data. *Biomarker Research*, 5(1).
- [3] arXiv:1901.05555
- [4] Wang, G. and Aban, I. (2015). Application of inverse probability weights in survival analysis. *Journal of Nuclear Cardiology*, 22(4), pp.611-613.
- [5] Huang, Z., et al. (2019). SALMON: Survival Analysis Learning With Multi-Omics Neural Networks on Breast Cancer. *Frontiers in Genetics*, 10.