INTEGRATION STRATEGIES FOR MULTI-OMICS SURVIVAL ANALYSIS

Francesca Calanca[1], Andrea Raiconi[1], Claudia Angelini[1]
francescacalanca@cnr.it
1 Istituto per le Applicazioni del Calcolo "Mauro Picone", Consiglio Nazionale delle Ricerche (CNR), 80131 Naples, Italy

BACKGROUND

Integrating high-dimensional, heterogeneous, potentially correlated **multi omics datasets** constitutes a **statistical challenge**, particularly for **survival outcomes**.

In this work, we benchmark several **supervised learning methods** aimed at estimating survival or hazard function, while providing also variable selection and patient score prediction. In particular, we consider different **data integration strategies** (illustrated in Figure *b*) aiming to extract hidden biological signals, improve the accuracy of survival models, and facilitate patient risk stratification as well as the discovery of novel biomarkers.

METHODS

Let $\{X^{(1)}, X^{(2)},..., X^{(K)}\}$ denote K omics datasets, each $X^{(k)} \in R^{p_k}$ for $k = 1,..., K$, represents the k-th omics block with p_k features. Associated with each sample are the right-censored survival data (Y_i, δ_i) *for* $i = 1,..., n$, where $Y_i = min(T_i, C_i)$, T_i being the observed survival time and C_i the censoring time; δ_i the event indicator (i.e., $\delta_i = 1$ if the event is observed, $\delta_i = 0$ if censored).

Our **workflow** (illustrated in Figure *a*) incorporates multiple preprocessing steps, including handling of missing values and sample intersection across omics blocks. Variable screening is applied to mitigate model overfitting in subsequent analyses. The pipeline supports both **single-omics and multi-omics modality**, offering a range of survival modeling approaches: Lasso, Adaptive Lasso, and Elastic Net penalized and Network-regularized Cox regression or AFT model; random survival forests; and a cooperative learning framework (Table 1). Hyperparameter tuning is performed via 5-fold cross-validation on the training set.

RESULTS

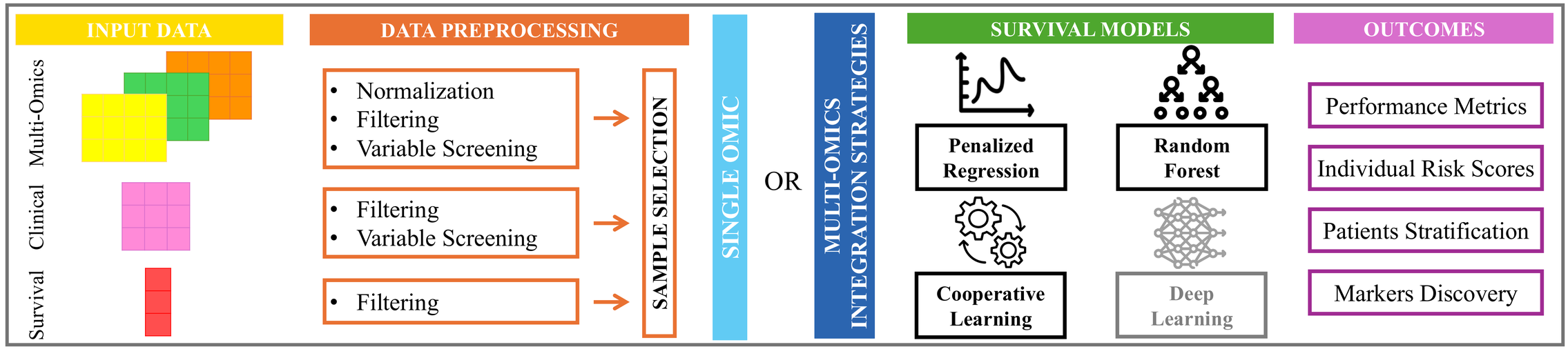
Trained models are evaluated on the test set using primarily C-index. Downstream analyses include **patient stratification** based on **predicted risk scores**, variable selection for **biomarker discovery**, and gene set **enrichment** for pathway-level interpretation.

The workflow is implemented within a **novel R PACKAGE Omics2Surv** that supports data acquisition from public repositories (e.g., TCGA via LinkedOmics).

Preliminary results from applications to multiple cancer cohorts suggest that models based on multi-omics integration often outperform single- omics models.

The package, datasets and results will be displayed in the accompanying smart presentation.

a) OVERALL WORKFLOW



b) MULTI-OMICS INTEGRATION STRATEGIES

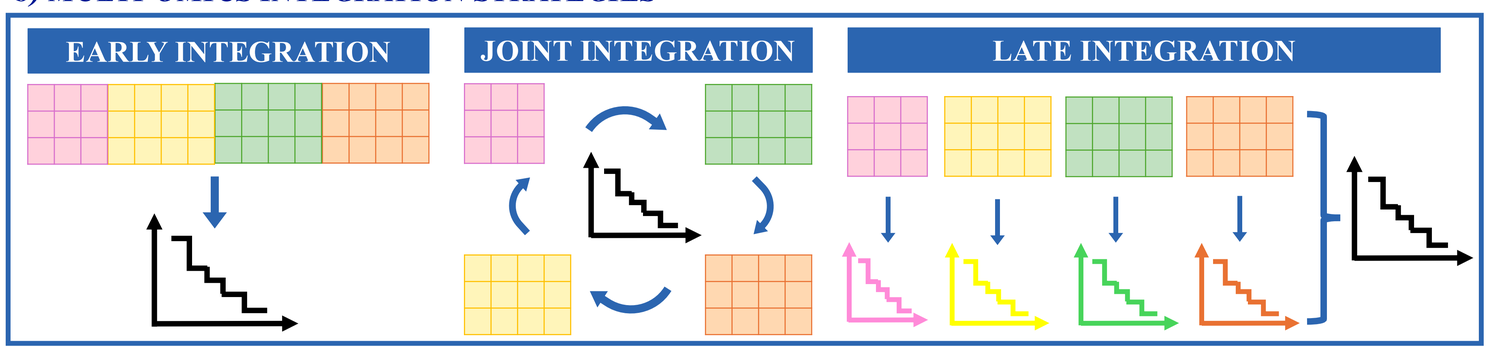


TABLE 1 Current Available Methods

	Survival Model	Package	Integration Strategy
SINGLE OMIC	Penalized Regression	`GLMNET [1]`, `COSMONET[2]`	
MULTI-OMICS	Penalized Regression	`GLMNET[1]`	**Early + Late**
	Random Forest	`BLOCKFOREST[3]`	**Joint**
	Cooperative Learning	`[4]`	**Joint**

Future Methods

Implementations of Deep Learning frameworks and AFT based methods [5] are ongoing.

[1] Simon, N., Friedman, J. H., Hastie, T., & Tibshirani, R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. Journal of Statistical Software, 39(5), 1–13. (2011)
[2] Iuliano, A., Occhipinti, A., Angelini, C., De Feis, I., & Liò, P. COSMONET: An R Package for Survival Analysis Using Screening-Network Methods. Mathematics, 9(24), 3262. (2021)
[3] Hornung, R., Wright, M.N. Block Forests: random forests for blocks of clinical and omics covariate data. BMC Bioinformatics 20, 358 (2019)
[4] Hahn, G., Prokopenko, D., Hecker, J., Lutz, S.M. et al. Prediction of disease-free survival for precision medicine using cooperative learning on multi-omic data. Briefings in Bioinformatics 25(4). (2024)
[5] Angelini, C., De Canditiis, D., De Feis, I. and Iuliano, A. A Network-Constrain Weibull AFT Model for Biomarkers Discovery. Biometrical Journal., 66: e202300272 (2024)

This work was supported by PRIN 2022 PNRR P2022BLN38 project, "Computational approaches for the integration of multi-omics data" funded by European Union - Next Generation EU, CUP B53D23027810001.