



# Ensemble Learning for Integrative Multi-Omics Modeling in Personalized Cancer Treatment

Jonathan Miller\*

Department of Biomedical Data Science, Massachusetts Institute of Technology, Cambridge, USA

## DESCRIPTION

Personalized cancer treatment represents a significant advancement in oncology, shifting the paradigm from traditional “one-size-fits-all” therapies to tailored interventions based on individual patient biology. At the heart of this transformation is the integration of multi-omics data genomic, transcriptomic, epigenomic, proteomic and metabolomic profiles that collectively capture the complex molecular landscape of tumors. However, the vast and heterogeneous nature of these datasets presents unique analytical challenges. Differences in scale, measurement techniques, missing data and biological variance between omics layers make data integration difficult. To address these challenges and enhance predictive accuracy, ensemble learning a machine learning strategy that combines multiple predictive models to form a more robust composite model has emerged as a powerful approach for integrative multi-omics modeling in personalized cancer treatment. Ensemble methods not only improve model generalizability and robustness but also help uncover subtle biological patterns and interdependencies that single-model approaches may miss, thereby enabling more accurate predictions of drug responses, patient stratification and clinical outcomes.

The foundational idea behind ensemble learning is simple yet powerful: rather than relying on a single algorithm to make predictions, it aggregates the strengths of multiple models. This can be achieved through various strategies such as bagging, boosting, stacking, or hybrid ensembles. In the context of multi-omics integration, each omics layer such as mutations (genomics), expression profiles (transcriptomics), protein abundance (proteomics), or metabolic signatures (metabolomics) can be analyzed using specialized base learners. These learners extract relevant features and patterns unique to that data type. Subsequently, their outputs are combined in a meta-learning step where a final model is trained to synthesize these predictions into a unified, clinically actionable result. This layered approach is particularly valuable in cancer research, where no single data source can fully explain tumor behavior or

therapeutic outcomes. A key advantage of ensemble learning in this setting is its ability to handle high-dimensional data with relatively few samples a common scenario in clinical datasets. Methods like Random Forests(RF) and Gradient Boosting Machines (GBM) inherently perform feature selection, reducing overfitting and identifying the most informative biomarkers from thousands of molecular variables. Furthermore, stacking methods allow the inclusion of diverse model types, such as decision trees, support vector machines and deep neural networks, each contributing unique insights from specific omics layers. For instance, while gene expression data may offer insights into pathway activation, proteomic data might highlight post-translational modifications relevant to drug resistance. Combining these views increases the fidelity and resolution of cancer characterization.

In personalized cancer treatment, ensemble learning has been used effectively for several key tasks. One is patient stratification, where patients are clustered into molecular subtypes that correlate with prognosis or treatment response. By integrating data from multiple omics layers, ensemble clustering methods like Similarity Network Fusion (SNF) have been used to identify robust and reproducible subgroups in cancers such as breast, colorectal and glioblastoma. These subtypes often go beyond classical histopathological classifications and reveal underlying molecular mechanisms, which can guide the use of targeted therapies or immunotherapies. Another important application is drug response prediction. Ensemble models trained on pharmacogenomic datasets like GDSC (Genomics of Drug Sensitivity in Cancer) and CCLE (Cancer Cell Line Encyclopedia) can be used to predict how a patient’s tumor will respond to specific compounds. These predictions are based on integrative profiles that include genetic mutations, expression levels, epigenetic modifications and protein targets. Such predictive models can help oncologists select the most effective treatment regimens, minimize trial-and-error prescribing and reduce adverse effects by avoiding ineffective therapies.

In terms of survival analysis, ensemble learning also offers significant benefits. Survival outcomes are inherently censored

**Correspondence to:** Jonathan Miller, Department of Biomedical Data Science, Massachusetts Institute of Technology, Cambridge, USA, Email: jmiller@mitdatasci.us

**Received:** 24-Feb-2025, Manuscript No. JDMGP-25-29286; **Editor assigned:** 26-Feb-2025, Pre QC No. JDMGP-25-29286 (PQ); **Reviewed:** 12-Mar-2025, QC No. JDMGP-25-29286; **Revised:** 18-Mar-2025, Manuscript No. JDMGP-25-29286 (R); **Published:** 26-Mar-2025, DOI: 1035248/2153-0602.25.16.370

**Citation:** Miller J (2025). Ensemble Learning for Integrative Multi-Omics Modeling in Personalized Cancer Treatment. J Data Mining Genomics Proteomics.16: 370.

**Copyright:** © 2025 Miller J. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

and time-dependent, which complicates standard regression techniques. Ensemble-based survival models like Random Survival Forests (RSF) or Coxboost incorporate omics-based features to predict survival probabilities over time. By leveraging multiple data types, these models improve risk stratification and allow clinicians to identify high-risk patients early, facilitating closer monitoring and more aggressive interventions. One of the major concerns in adopting machine learning, particularly in clinical environments, is model interpretability. Clinicians and researchers need to understand not only the predictions but also

the rationale behind them. Ensemble learning models, especially tree-based methods like Random Forests or GBM, offer mechanisms to assess feature importance, helping identify which genes, proteins, or metabolites most strongly influence a given outcome. This not only aids biological interpretation but also accelerates the discovery of novel biomarkers and therapeutic targets. Additionally, newer approaches in explainable AI, such as SHAP (Shapley Additive Explanations), can be used alongside ensemble models to provide transparent, case-specific explanations of model behavior.