Cancer survival prediction and integration of multi-omics integration with Supervised Autoencoders, Stacked Autoencoders and Concrete Supervised Autoencoders for multiple correlated driver genes

Vasileios Alevizos
Karolinska Institutet,
Stockholm, Sweden
ORCID: 0000-0002-3651-2134

Yishu Qu
Northwestern University,
Illinois, USA
YishuQu2024@u.northwestern.edu

Alexus Acton
University of Alabama,
Birmingham, USA
actonam@uab.edu

Vanessa Xiao
Massachusetts Institute of Technology
Cambridge, USA
vzxiao@mit.edu

Zongliang Yue
University of Alabama,
Birmingham, USA
zongyue@uab.edu

Abstract: In recent years, personalized treatment approaches to breast cancer have emerged in a progressive number of patients and mortality rate worldwide. Multidisciplinary studies have been regarded as an irrefutable fact in literature, as standardization of the procedure became dubious. Omics approaches endeavor precise treatment during therapy ameliorating deadly dysfunction. Thus, integration of accumulated multi-omics might elaborate on cancer survival scaling up manifold omics analysis and cancer subtyping. In this study, we evaluate robustness of autoencoders extracting subtype features and accuracy predicting breast cancer. Deep learning models revealed reliable cancer prognosis in respect to performance.

Keywords—breast cancer, multi-omics, autoencoder, cancer subtyping.

Background: Breast cancer is a heterogeneous disease that differs among patients as well as intratumorally. With the application of single-layer omics utilizing mRNA expression, breast cancer has been characterized to have substantial differences in both the pathological and molecular characteristics that drive tumorigenesis. These deep learning methods have advanced the understanding of breast cancer allowing for molecular subtyping [2]. Currently, understanding the molecular and cellular characteristics of tumor heterogeneity that is relevant to the diagnosis, prognosis, and standard of care of breast cancer is essential for improving patient quality of life.

Methods: *Denoising Autoencoder for accurate cancer prognosis prediction* (DCAP) is a framework that allows for the integration of multi-omics data by denoising. The unsupervised denoising autoencoder allows for high dimensional input data that is utilized to accurately estimate cancer risks through the Cox model [1].

eXtreme Gradient Boosting (XGboost) is an algorithm used for the implementation of gradient boosted tree algorithm. This supervised learning algorithm allows for the prediction of target variables by combining estimates of a set of other models. XGBoost allows for the selection of a small number of genomic features that are considered to be correlated to tumorigenesis and risk factor level of breast cancer [1].

Multi-Omics Autoencoder-based Neural Networks Algorithm (Moanna) is a deep-learning-based algorithm that utilizes a semi-supervised autoencoder layer jointly trained with supervised feed-forward neural network multi-task classification layers. It integrates multi-omics data, in particular copy number and somatic mutations, for predicting breast cancer subtypes and hormone receptor status [3].

Conclusion and future work: This project field the need for more robust cancer subtype diagnosis using deep learning methods such as DCAP (A framework to integrate multi-omics data by Denoising Autoencoder for Accurate cancer prognosis prediction) concluding mRNA performance better in comparison to miRNA, methylation and following CNV (Copy Number Variation). These constructed models could distinguish high-risk patients from low-risk by identifying at the same time breast cancer related biomarkers. Room for performance and improvements still exists while these empirical results might be utilized to impact hitherto patients considering GAN (Generative Adversarial Network) architectures by prioritizing synthetic data corresponding to important omics profiles [4].

Acknowledgement: Authors feel obliged to thank all the instructors and organizers from the UAB Informatics Institute for hosting the undisputable successful hacking omics event. This academic contribution was under the aegis of the University of Alabama at Birmingham.

References (APA style):

[1] Chai, H., Zhou, X., Zhang, Z., Rao, J., Zhao, H., & Yang, Y. (2021). Integrating multi-omics data through deep learning for accurate cancer prognosis prediction. Computers in biology and medicine, 134, 104481.

[2] Wang, X., Hu, Z., Yu, T., Wang, R., Wei, Y., Shu, J., ... & Li, Y. (2021). Contrastive cycle adversarial autoencoders for single-cell multi-omics alignment and integration. arXiv preprint arXiv:2112.03266.

[3] Lupat, R. (2020). Automated Discovery of Interacting Genomic Events That Impact Cancer Survival by Using Data Mining and Machine Learning Techniques.

[4] Ahmed, K. T., Sun, J., Cheng, S., Yong, J., & Zhang, W. (2022). Multi-omics data integration by generative adversarial network. Bioinformatics, 38(1), 179-186.