DIANE: A Novel Multi-Omics Data Integration Framework Using Attributed Network Embedding Paired With Machine Learning for More Accurate Biomedical Classification

Chaturvedi, Amogh (School: Canyon Crest Academy)

Advancements in high-throughput technologies have expanded omics repositories, which has created the need to develop novel computational methods. Currently, data integration methods are limited as they are computationally expensive, prone to overfitting on small datasets, and cannot work effectively with missing data. This project presents a novel, generalized framework called Data Integration using Attributed Network Embedding (DIANE) that exploits high-dimensional multi-omics data for more accurate biomedical classification. DIANE, consisting of three steps, starts with preprocessing which includes a novel multi-omics feature selection pipeline that aids in dimensionality reduction. Then, using bivariate correlation methods, a patient similarity network is constructed, which in addition to the multi-omics signatures, will be fed into modified attributed network embedding algorithms (ANEs) to produce a latent representation. Finally, the latent representation will be used to train supervised machine learning classifiers. We assess the performance and prove the generalizability of DIANE by testing it on two breast cancer classification problems: A multi-class PAM50 subtype classification & a binary classification between two histological subtypes (Invasive Ductal Carcinoma & Invasive Lobular Carcinoma). For the two classification problems, DIANE outperforms popular state-of-the-art data integration methods both in performance and in minimizing overfitting. Furthermore, DIANE proves to be robust with missing data and flexible in working with different numbers of omics. This study has many innovations: DIANE is the first data integration method to use unsupervised ANEs and the first to apply a multi-omics approach for invasive ductal and lobular carcinoma classification.

Awards Won:

Third Award of $1,000

Third Award of $1,000

Lawrence Technological University: Tuition scholarship of $19,650 per year, renewable for up to four years and applicable to any major

American Statistical Association: In-Kind membership to ASA for all winners, including honorable mentions

American Statistical Association: Honorable Mention