DAE-XGBOOST: Developing a Method for Classifying Neoplasms via MicroRNAs for Biopsy Verification

Hu, Emily (School: Paul Laurence Dunbar High School)

With respect to evaluating biopsies for cancer diagnosis, a pressing yet often sidelined issue is that of anatomical pathology errors. Due to the image-based nature of biopsy examinations and the variation between cell appearances, the current process for verifying and diagnosing cancers has been characterized with a high level of inconsistencies between clinical opinions, and at times, incorrect diagnosis. In a substantial number of these cases, pathological errors have resulted in detriments to patient treatment outcomes. As such, this project sought to develop an efficient and robust molecular framework to evaluate biopsies in a way that is safeguarded against pathological mistakes caused by human error. Specifically, dimensional compression by denoising autoencoder (DAE) neural networks and tree-based ensemble classification via XG-Boost, were utilized to both learn the key microRNA features of each neoplasm, and then classify the tumors based on the molecular features. microRNAs were selected as the target molecule due to their short length and role in regulating cell growth and differentiation. Classification between all neoplasms demonstrated an overall accuracy of 94.01%, and the binary classifications between tumorous cells and normal cells was highly discriminant, with many neoplasm types correctly classified at close to 100%. The high sensitivity between all neoplasms, and more importantly, between normal and neoplastic cells based on their miRNA profile presents a promising verification method for biopsies. If applied in a clinical setting, this computational-molecular framework has the potential to serve as a cost-effective baseline for biopsy purity and validation for pathologists, improving patient treatment outcomes.