[BIOINFO 2024]

Leveraging machine learning on RNA-seq data for early discrimination of Alzheimer's Disease from cognitively unimpaired individuals

Hyun Woo Park[1], Ki Wook Lee[1], Jimin Seo[1], Na Young Kwon[1], Hye Jung Min[1], Heeje Cho[1],

Minho Choi[1], Balachandran Manavalan[2], and Young-Jun Jeon[1,*]

[1] *RNA-based Precision Medicine Laboratory, Department of Integrative Biotechnology,*
Sungkyunkwan University
[2] *Computational Biology and Bioinformatics Laboratory, Department of Integrative Biotechnology,*
Sungkyunkwan University
* *Corresponding author: jeon2020@skku.edu*

Alzheimer's disease (AD) is a progressive neurodegenerative disorder with limited early diagnostic options, primarily due to the absence of reliable biomarkers. Early intervention is often delayed, as current diagnostic tools, like the Mini-Mental State Examination (MMSE), rely heavily on subjective assessments, highlighting the need for objective and accurate biomarkers. Recent advancements in RNA sequencing (RNA-seq) offer a promising avenue for exploring transcriptomic alterations associated with AD at the molecular level. This study utilizes RNA-seq data from brain tissues to distinguish AD patients from cognitively unimpaired (CU) individuals using machine learning techniques. By analyzing gene expression profiles, we aim to identify diagnostic signatures sensitive and specific to early stages of AD. Traditional machine learning algorithms are applied to the RNA-seq datasets to detect patterns linked to AD pathology. Our findings demonstrate that machine learning can effectively classify AD patients, with metrics such as area under the ROC curve (AUROC), Matthew's Correlation Coefficient (MCC), accuracy, sensitivity, and specificity validating RNA-seq as a powerful tool for detection. Further optimization significantly enhances model performance. This approach lays the groundwork for predictive models that could facilitate earlier, more accurate diagnosis and treatment of Alzheimer's disease.