Child Health Research Institute

University of Nebraska Medical Center

Children's NEBRASKA

Integrating Multi-Omics Data by a Multi-Modal Transfer Learning Model to Reduce Healthcare Disparities for Kidney Renal Clear Cell Carcinoma

Lusheng Li[1] and Shibiao Wan[1]
[1]Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, NE 68198

Kidney cancer is among the 10 most frequently diagnosed cancers in the United States. It is estimated that 81,800 cases will be diagnosed kidney cancer in 2023, with 14,890 deaths. Kidney Renal Clear Cell Carcinoma (KIRC) is the most common subtype of kidney cancer, accounting for approximately 80% of all cases. Although significant progress has been made to improve KIRC prognosis, existing studies indicated that Black American have disproportionately high incidence and mortality rates in kidney cancer compared with White American. With artificial intelligence (AI) being increasingly applied to KIRC research and clinical decision-making, KIRC data disparities would introduce bias to AI/ML models and further enhance negative impacts on healthcare towards underrepresented groups. Transfer learning has shown potential to reduce racial disparities in KIRC. However, its performance may be deteriorated due to the following disadvantages: its ML model requires large-scale training samples which are difficult to obtain in clinical settings, and it only uses single-omics data without integrating multi-omics information. To address these concerns, we propose to develop a multi-modal transfer learning model to integrate multi-omics data for reducing health disparities. Specifically, we first investigated two multi-modal ensemble methods, Pearson Correlation Coefficient (PCC) based patient-pairwise similarity, and variational autoencoder (VAE) to integrate different omics data. Then, we leveraged a transfer learning model based on domain adaptation to pre-train the model on the majority group (White Americans) and fine-tune the model using the minority group (Black Americans). To further address the imbalanced data among ethnic groups, we explored implementing a data augmentation method, Synthetic Minority Oversampling Technique (SMOTE), to increase the minority group data. We evaluated our model on multi-omics data (mRNA, miRNA and methylation) and clinical outcome endpoints of KIRC from the TCGA database. Experimental results suggested that our proposed approach achieved better performance of reducing health disparities for Black Americans compared with the mixture model and the independent model as well as the conventional transfer learning model in classifying overall survival (OS) and progression-free interval (PFI) prognosis category. Furthermore, SMOTE can help alleviate the imbalanced problem and improve the performance of the prognosis classification for ethnic minority groups. Our findings demonstrate that our proposed multi-modal transfer learning approach effectively reduces the performance gap between majority and minority groups in KIRC prognosis, thereby mitigating healthcare disparities. The integration of multi-omics data and the implementation of data-augmentation methods like SMOTE hold promise for more equitable healthcare solutions with AI-driven biomedicine.