

Dear Editor,

We would like to submit our work "Attention-based pattern discovery of mass spectrometry imaging data" to Advanced Science.

Mass spectrometry imaging (MSI) enables the direct visualization of molecular distributions in tissue sections, making it a crucial method in metabolomics research. However, the vast size and high dimensionality of MSI data pose challenges for analysis even though there are already many conventional machine learning methods used in this area, such as the "Curse of dimensionality" problem. Therefore, compressing sparse distributions of raw data while ensuring minimal information loss is important. In response to these challenges, we propose Atnal, an attention-based generative model. Atnal effectively maps MSI data to a low-dimensional space with an extremely low loss ($2 \times 10^{-7} \sim 7 \times 10^{-9}$), which can contribute to the pattern discovery of MSI data. Then Atnal is applied in the domain of cancer region recognition and correlation analysis, based on the metabolomics data collected from patients with prostate cancer and colorectal adenocarcinoma. As it is presented, Atnal can distinguish the regions primarily containing cancer cells from those with normal cells and identify highly correlated metabolites with cancer (correlation coefficient up to 0.79). Atnal can provide quantitative guidance for the clinical removal of cancerous tissue, helping to avoid subjective bias and further aid in clinical cancer diagnosis.

Thank you for your attention!

Sincerely yours,

Cong Wu