



XXXIII SIC SALÃO INICIAÇÃO CIENTÍFICA

Evento	Salão UFRGS 2021: SIC - XXXIII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2021
Local	Virtual
Título	Graph Attention Networks para a integração de dados ômicos e classificação de estadiamento tumoral
Autor	FELIPE COLOMBELLI
Orientador	MARIANA RECAMONDE MENDOZA GUERREIRO

Graph Attention Networks para a integração de dados ômicos e classificação de estadiamento tumoral

Com os avanços do poder computacional e as tecnologias de sequenciamento de nova geração, dados ômicos têm recebido crescente atenção. Dados de expressão gênica, por exemplo, são amplamente utilizados devido ao seu alto grau informativo e potencial para responder uma ampla gama de questões biológicas, como a diferenciação entre perfis transcriptômicos de pacientes saudáveis e pacientes com câncer. Dados de expressão de microRNA e dados de metilação de DNA também estão se popularizando, dando surgimento a importantes questões como a integração destes e outros dados ômicos, cunhando o termo multi-ômicas. Além disso, apesar da importância da tarefa de classificação de estadiamento tumoral e da ampla disponibilidade desses dados ômicos, os esforços nessa direção ainda são limitados. Frente a esta difícil tarefa de classificação e visando contribuir para questões como a integração de multi-ômicas, buscamos investigar abordagens baseadas em aprendizado profundo para grafos, em particular a Graph Attention Network (GAT), para classificação de estadiamento tumoral de câncer renal, além de comparar como este modelo difere de abordagens tradicionais, especificamente a Multilayer Perceptron (MLP). Para a MLP, a integração dos dados foi feita de maneira direta, concatenado as ômicas em uma única grande tabela. Para a GAT, foram geradas 5 redes de pacientes, calculando suas correlações sob os dados de expressão gênica e ligando-os de acordo com valores mínimos do coeficiente de correlação de Pearson absoluto e de um p-valor máximo de 0.05. O restante das ômicas foram integradas como atributos dos nós das redes. Utilizando ou não a integração de ômicas, a GAT foi destaque, superando a MLP. Em todos os experimentos, atingiu a maior PR-AUC (0.654) no primeiro cenário e a maior ROC-AUC (0.845) no segundo. O excelente desempenho com múltiplas ômicas, caracterizado por um tamanho amostral duas vezes menor, sugere grande potencial da abordagem na resolução deste problema.