

**PREDIÇÃO DOS LOCAIS DE CLIVAGEM DA PROTEASE DO
HIV-1**

IBI5031 - RECONHECIMENTO DE PADRÕES I

Maycown Douglas de Oliveira Miranda

1. DESCRIÇÃO

As proteases são enzimas que atuam na quebra de ligações peptídicas das proteínas. A clivagem proteolítica, como é denominado este processo, é a quebra destas ligações entre os aminoácidos. [1]

O vírus HIV, em específico o HIV-1, possui uma necessidade essencial da protease para seu ciclo de vida. Sendo assim, um modelo preditivo para a previsão da clivagem pela protease pode ajudar no desenvolvimento de melhores e mais potentes inibidores de protease, e com menos efeitos colaterais [2]

Na literatura são apresentadas duas abordagens para a predição da clivagem: modelagem molecular e análise de sequências dos aminoácidos. Esta última tem sido bastante explorada, possuindo mais trabalhos relevantes [3]. Dentre estes a utilização de Cadeia de Markov [4], algoritmos genéticos para seleção de características [5], utilização de *linear support vector machines* (LSVM) com codificação ortogonal padrão [3].

2. MÉTODOS

Este trabalho pretende explorar a aplicação de algoritmos de redes neurais recorrentes nas bases utilizadas no trabalho de [3] a fim de verificar a performance e aplicabilidade, demonstrando se esta pode ser uma abordagem a ser utilizada para melhorar a classificação realizada em [3].

3. RESULTADOS ESPERADOS

Espera-se obter um valor de AUC maior ou igual ao do trabalho de [3] com a metodologia de avaliação dos autores, efetuando o treinamento em uma das bases, e verificando o valor de AUC nas demais. Busca-se também, avaliar demais métricas como *F1-Score*, *Matthews correlation coefficient*, acurácia e *AUCPR*.

Caso o valor de AUC encontrado seja menor, será uma evidência de que tal modelagem não se demonstrou eficiente neste tipo de problema, e que uma maneira diferente de codificação e pré-processamento seja necessária.

Entretanto, caso o valor seja maior ou igual, ter-se-á sucedido a modelagem.

4. CONCLUSÕES

Este projeto poderá contribuir para a exploração de novas metodologias, bem como a mensuração da performance da modelagem em relação aos resultados mais promissores da literatura.

5. BIBLIOGRAFIA

- [1] A. J. Barret, N. D. Rawlings e J. F. Woessner, *The Handbook of Proteolytic Enzymes*, San Diego: Elsevier, 2004.
- [2] E. Devroe, P. A. Silver e A. Engelman, "HIV-1 incorporates and proteolytically processes human NDR1 and NDR2 serine-threonine kinases," *Virology*, pp. 181-189, 2005.
- [3] T. Rögnvaldsson, L. You e D. Garwicz, "State of the art prediction of HIV-1 protease cleavage sites," *Bioinformatics*, pp. 1204-1210, 2015.
- [4] H. Oğul, "Variable context Markov chains for HIV protease cleavage site prediction," *Biosystems*, pp. 246-250, 2009.
- [5] B. Niu, X.-C. Yuan, P. Roeper, Q. Su, C.-R. Peng, J.-Y. Yin, J. Ding, H. Li e W.-C. Lu, "HIV-1 Protease Cleavage Site Prediction Based on Two-Stage Feature Selection Method," *Protein & Peptide Letters*, pp. 290-298, 2013.
- [6] <https://archive.ics.uci.edu/ml/datasets/HIV-1+protease+cleavage#>
- [7] <https://github.com/czbiohub/learn-bioinformatics>
- [8] <https://www.helsinki.fi/en/news/data-science-news/a-new-direction-in-evolutionary-research-predicting-the-evolution-of-viruses-bacteria-and-cancer-cells>