

Giordano Bruno Olivetti Mattiello - 173056

Daniela Marques de Moraes - 169562

Predição de Efetividade do tratamento de HIV

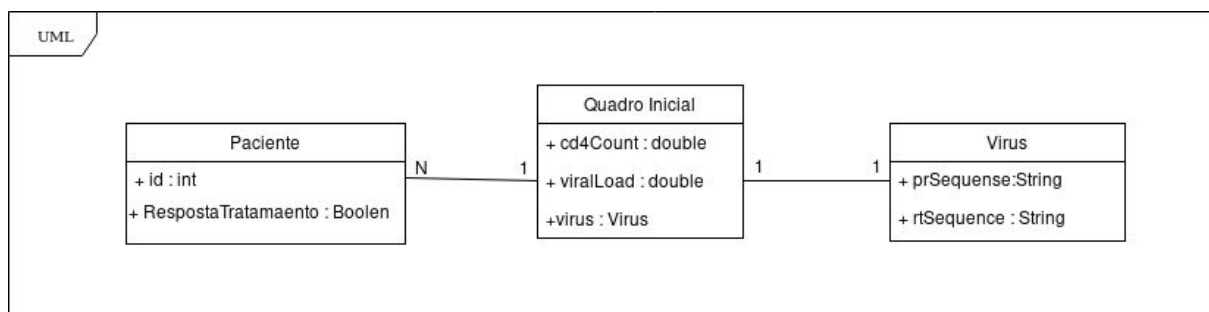
O vírus HIV tem matado milhões de pessoas ao redor do mundo todo, foi reconhecido em 1981 pela primeira vez desde então muito esforços foram destinados a descobrir o seu funcionamento e seus tratamentos.

Um passo para a compreensão do vírus é entender o seu projeto genético, neste estudo pretendemos identificar a relação entre a estrutura do vírus e a efetividade do seu tratamento usando machine learning para criar estes resultados.

O Predict HIV Progression (<https://www.kaggle.com/c/hivprogression/data>) é uma base de dados que utilizou de Machine Learning para prever o progresso de pacientes portadores de HIV após 16 semanas de tratamento, a qual conseguiu especificar a sequência de nucleótidos e transcriptase nucleotídios dos vírus.

Para conseguirmos relacionar a formação genética dos vírus com a efetividade do tratamento, modelamos o banco de dados da seguinte forma: Foram modelados 3 tabelas, como a tabela vírus, quadro do paciente e paciente. Para a tabela “vírus” é necessário as sequências dos vírus (Transcriptase Reversa e Protease). Em o quadro do paciente, consideramos que uma melhora efetiva é a diminuição do atributo *ViralLoad* em até 100 vezes (representada na escala log). Portanto, modelamos essa tabela para conter os atributos *viralLoal*, *cd4Count* e a relação com a tabela vírus. Por fim, na tabela Paciente temos algumas informações como o ID e o status do seu tratamento (se foi efetivo ou não)

Modelo UML



A utilização da sequência nucleotídica da Transcriptase Reversa (RT) e Protease (PR) para prever a progressão a curto prazo do paciente. Em modo geral, a sequência de nucleotídeos é a identidade da proteína, que é o sistema de defesa da célula. A enzima RT é responsável por copiar o genoma do HIV-1 dentro da célula.

Como o genoma do HIV-1 é traduzido, está em uma longa sequência de aminoácidos; a proteína PR corta essa cadeia nas numerosas unidades funcionais - exigidas pelo ciclo de vida do HIV. Estas são as proteínas que são alvo da maioria dos medicamentos para o HIV-1, uma vez que são principalmente exclusivas do ciclo de vida do HIV-1.

Juntamente com as sequências virais do HIV-1, são fornecidos os dois indicadores clínicos comuns usados para determinar a "saúde geral" de um indivíduo infectado pelo HIV-1: Carga Viral e contagem de células CD4 +. A contagem de células CD4 + é uma estimativa do número de glóbulos brancos em 1 mL de sangue, enquanto a carga viral é o número de partículas virais nesse mesmo mL. Neste conjunto de dados, a carga viral é representada em uma escala log-10.