

SUBMISSÃO / RESUMO-EDITAIS (PROPEGI/PROEC/PROGRAD) -
TRABALHOS LOCAIS - PESQUISA

**PREDIÇÃO DA PROGRESSÃO NEUROLÓGICA EM PESSOAS VIVENDO
COM HTLV UTILIZANDO MODELOS DE MACHINE LEARNING**

Laryssa Bandeira De Melo Silva (laryssa.bandeira@upe.br)

Patrícia Muniz Mendes Freire De Moura (patricia.moura@upe.br)

João Pacífico Bezerra Neto (pacifico.joao@upe.br)

O vírus linfotrópico de células T humanas (HTLV) pode causar desde infecções assintomáticas até doenças graves, como a paraparesia espástica tropical (HAM/TSP). A ausência de biomarcadores específicos dificulta prever a progressão neurológica. Nesse cenário, a aplicação de machine learning (ML) surge como alternativa promissora, integrando grandes volumes de dados para identificar padrões ocultos. Este estudo utilizou modelos de ML para prever sintomas neurológicos em pessoas vivendo com HTLV, com base em dados clínicos e hematológicos. Foram incluídos 127 pacientes acompanhados no Hospital Universitário Oswaldo Cruz. O projeto foi aprovado pelo Comitê de Ética em Pesquisa do HUOC (CEP-HUOC), sob parecer nº 5.543.999. As variáveis foram organizadas em dataframes e pré-processadas, incluindo eliminação de inconsistências, normalização de valores numéricos e conversão

de variáveis categóricas. A relação entre elas foi avaliada pelos métodos de Spearman e t-student, seguida da seleção de atributos pelo SelectKBest e remoção de multicolinearidade via VIF. Para treino e teste aplicou-se validação cruzada Leave-One-Out combinada com random under sampler. Os dados foram submetidos a nove modelos de ML e avaliados por recall, curva de aprendizado, matriz de confusão e importância das variáveis (SHAP). Os modelos Random Forest (RF) e Logistic Regression (LR) apresentaram melhor desempenho, com boa generalização e baixa taxa de falsos negativos (5,5% e 6,3%, respectivamente), essencial para detecção precoce do risco neurológico. Embora o Naive Bayes tenha alcançado recall superior (0,812) e apenas 5 falsos negativos, mostrou tendência ao overfitting pela divergência entre treino e teste. A análise SHAP apontou doenças cardiovasculares como principais preditores (+0,19 na RF; +0,51 no LR), seguidas por sintomas gastrointestinais e leucócitos. Assim, modelos de ML mostram-se promissores para prever a progressão clínica, apoiando a prática médica e o prognóstico personalizado.

Palavras-chave: variáveis; progressão; modelos.