MACHINE LEARNING PARA DETECÇÃO DE PADRÕES DE CLASSIFICAÇÃO EM DADOS DE ANAMNESE E DE ANÁLISE CITOPATOLÓGICA PARA DETECÇÃO DE CÂNCER DE COLO DO ÚTERO

Otávio Luís Dal Molin Folador (PIBIC/CNPq), Andréa Timóteo dos Santos, Ednéia Peres Machado, Alaine Margarete Guimarães (Orientadora), e-mail: alainemg@uepg.br.

Universidade Estadual de Ponta Grossa / Departamento de Informática.

Ciência da Computação: Metodologia e Técnicas da Computação.

Palavras-Chave: *Aprendizado de Máquina, Câncer Cervical, Exame de Papanicolaou, Features Selection, Redes Neurais.*

Resumo

Este trabalho buscou analisar a correlação entre atributos relevantes de dados de anamneses com laudos de exames citológicos por meio de algoritmos de Aprendizado de Máquina (AM). Para avaliar a relevância dos atributos, foram utilizados algoritmos de Seleção de Atributos, pelos quais foram criados três conjuntos de dados. A partir desses dados, foram avaliados quatro modelos de AM a partir da precisão de suas previsões. Os resultados evidenciam que a redução no número de atributos favorece o desempenho dos modelos.

Introdução

No Brasil, o principal método de rastreio do câncer do colo do útero é o exame citológico, que coleta amostras de células do colo do útero para análise laboratorial. Essa coleta é acompanhada por uma anamnese clínica, na qual se registram antecedentes pessoais e familiares, que indicam potenciais fatores de risco. Os dados de anamneses e laudos citológicos podem ser analisados com técnicas de Aprendizado de Máquina (AM) a fim de explorar a correlação entre fatores de risco e o resultado do exame. Este estudo visou aplicar AM a uma base de dados de anamneses levantada na Clínica Solidária da Rede Feminina de Combate ao Câncer – Regional de Ponta Grossa (RFCC-PG) e no Laboratório Universitário de Análises Clínicas (LUAC) da Universidade Estadual de Ponta Grossa (Folador et al., 2024) a fim de avaliar a correlação de atributos relevantes nos dados de anamnese com o laudo de exames citológicos.

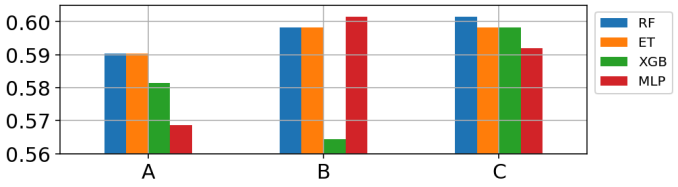
Materiais e Métodos

A base de dados passou por múltiplos algoritmos de Seleção de Atributos (SA), que são responsáveis por selecionar apenas um subconjunto dos atributos mais relevantes. A partir dessa técnica, foram produzidas 3 conjuntos de dados: (A) com todos os 833 atributos; (B) com os atributos que apareciam pelo menos uma vez entre os 30 atributos mais relevantes (92 atributos); (C) com os atributos que apareciam pelo menos uma vez entre os 10 atributos mais relevantes (36 atributos). A esses três conjuntos de dados foram aplicados os algoritmos de AM: Random Forest (RF), Extra Tree (ET), eXtreme Gradient Boosting (XGB) e MultiLayer Perceptron (MLP).

Resultados e Discussão

Os modelos de AM obtiveram melhores resultados para os conjuntos de dados que possuíam menores números de atributos. Os modelos de MLP com o conjunto B e RF com o conjunto C obtiveram a melhor pontuação de precisão, de 60.14%, como mostra a Figura 1.

Figura 1. Resultados dos modelos para as três bases de dados.



Observa-se que o modelo de RF obteve resultado equivalente ao modelo de MLP com menos atributos, o que torna seu processamento menos custoso computacionalmente.

Conclusão

A partir dos resultados obtidos, observa-se que a redução no número de atributos por algoritmos de SA apresentou melhor desempenho nos modelos de AM. Assim, pesquisas futuras poderão avaliar outras técnicas de AM para obter modelos capazes de realizar previsões com maior precisão, tomando como base o conhecimento de que a análise dos dados de anamnese é beneficiada pela seleção de atributos relevantes.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, ao LUAC e à RFCC-PG.

Referências

INSTITUTO NACIONAL DE CÂNCER (Brasil). Câncer. Tipos de câncer. **Câncer do Colo do Útero**. Rio de Janeiro: INCA, 2022. Disponível em: https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/colo-do-utero. Acesso em: 12 set. 2025.

FOLADOR, O. et al. Avaliação do Impacto do Uso de Técnicas de Feature Selection em Dados Relacionados ao Câncer Cervical. **Anais do 33º Encontro Anual de Iniciação Científica**, Ponta Grossa, nov. 2024. Disponível em: https://siseve.apps.uepg.br/storage/eaic2024/8_Otavio_Luis_Dal_Molin_Folador-172953518388987.pdf. Acesso em: 12 set. 2025.