

UTILIZAÇÃO DE MACHINE LEARNING PARA O DIAGNÓSTICO DE LEUCEMIA A PARTIR DE DADOS DE EXPRESSÃO GÊNICA

Joaquim dos Santos Costa¹; Maria Eliana da Silva Holanda²; Edson Magalhaes da Costa³; Jakelyne Machado Lima Silva⁴; Gilberto Nerino de Souza Junior⁵
Marcus de Barros Braga⁶

1. Bolsista PIBITI, Graduando Sistemas de Informação, Universidade Federal Rural da Amazônia, Campus Paragominas, e-mail: joaquimsants24@gmail.com; 2. Bolsista FAPESPA, Graduanda em Sistemas de Informação, Universidade Federal Rural da Amazônia, Campus Paragominas, e-mail: eliana.holanda85@gmail.com; 3. Graduando em Sistemas de Informação, Universidade Federal Rural da Amazônia, Campus Paragominas, e-mail: edsonmagalhaesdacosta@gmail.com; 4. Docente Co-orientadora, Universidade Federal Rural da Amazônia, Campus Paragominas, e-mail: jakelyne.silva@ufra.edu.br; 5. Docente Co-orientador, Universidade Federal Rural da Amazônia, Campus Paragominas, e-mail: gilbertojr@ufra.edu.br; 6. Orientador, Núcleo de Pesquisas em Computação Aplicada, Campus Paragominas, Universidade Federal Rural da Amazônia, e-mail: marcus.braga@ufra.edu.br.

RESUMO:

O grande desafio do tratamento e controle do câncer está no diagnóstico precoce, pois direcionar terapias corretas para seu controle, maximiza a eficácia e minimiza as rejeições no tratamento a longo prazo. A AML (*Acute Myeloid Leukemia*), ou leucemia mieloide aguda, é um câncer de sangue e da medula óssea onde há excesso de glóbulos brancos imaturos, e proliferação de células patologicamente alteradas na corrente sanguínea. Os sintomas podem incluir fadiga, febre, infecções recorrentes e aparecimento de hematomas. A ALL (*Acute Lymphoblastic Leukemia*), ou leucemia linfoblástica aguda, também afeta os glóbulos brancos e ocorre quando surge um linfócito imaturo e danificado na medula óssea devido a um erro no DNA. Os sintomas incluem aumento dos gânglios linfáticos, hematomas, febre, dor óssea, sangramento gengival e infecções. Apesar da distinção bem estabelecida entre a AML e a ALL, não há um método eficaz de teste que estabeleça um diagnóstico preciso. Para isso, a prática clínica envolve a expertise de um hematologista para interpretar a morfologia do tumor, a partir de uma série de exames realizados por laboratórios especializados. Deste modo, desenvolver um método de teste eficaz que firme um diagnóstico preciso, otimizará significativamente o tratamento dos tipos de leucemia e suas especificidades. A técnica de *microarrays*, ganhou visibilidade com a popularização das tecnologias de sequenciamento de genomas. Através desta técnica, ainda considerada como padrão ouro, é possível investigar milhares de genes simultaneamente. Nos últimos anos, estudos de expressão gênica têm sido utilizados na oncologia para identificar marcadores moleculares que possam diferenciar perfis de expressão e auxiliar no diagnóstico, classificação e escolha correta do tratamento. Por meio da determinação da expressão de milhares de genes simultaneamente, a técnica de *microarrays* possibilita realizar a comparação do comportamento molecular de vários tipos de linhagens celulares e tecidos diferentes, quando expostos a uma condição específica patológica ou experimental. Em 1999, Golub e colaboradores propuseram uma abordagem genérica para classificação do câncer com base no monitoramento da expressão gênica por DNA *Microarrays* e aplicaram a leucemias agudas humanas. Foi possível então se estabelecer um procedimento de classificação automática entre AML e ALL. Este projeto utilizou dados de expressão gênica e aplicou dois modelos de aprendizado de máquina (*machine learning*), uma subárea da inteligência artificial, para estabelecer o diagnóstico dos tipos de leucemia (ALL ou AML) a partir dos dados de expressão gênica dos pacientes estudados. Dois modelos foram utilizados, uma máquina de vetor de suporte (*support vector machine*, SVM) e uma rede neural artificial (*artificial neural network*, ANN). O SVM conseguiu uma acurácia de classificação de 67% e por isso, outro algoritmo de aprendizado de máquina, o ANN, foi testado. A rede neural obteve um resultado melhor, uma acurácia na classificação de 85%. Futuros trabalhos podem combinar vários algoritmos de inteligentes, abordagem conhecida como *machine learning ensemble*, para aprimorar o desempenho na classificação destes tipos de leucemia, porém, a abordagem de aprendizado automático já se mostrou capaz de realizar a tarefa do diagnóstico sem a necessidade de intervenção de expertise humana, normalmente requerida na abordagem médica tradicional.

PALAVRAS-CHAVE: leucemia; machine learning; expressão gênica.

¹ Link do Vídeo: <https://youtu.be/kl5OGFYONck>