

SOFTWARE · UECE · 30 DE DEZEMBRO DE 2025

One-Class Genetic Algorithm for Authentication Analysis of Spectrochemical Data

Por José R. de Moraes Filho, Camilo de L. M. de Moraes, Anne B. F. Câmara e Kássio M. G. Lima

Fonte: OpenAlex · UECE · [Publicação original](#)

INOVAÇÃO	APLICABILIDADE	POTENCIAL ECONÔMICO	MATURIDADE
8/10	9/10	9/10	Média

O estudo apresenta um algoritmo genético de classe única (OGA) para seleção de variáveis, combinado com classificadores de classe única (DD-SIMCA e OCPLS), demonstrando alta sensibilidade (100%) na análise de dados espectroquímicos para diagnóstico de COVID-19, endometriose e dengue. A abordagem oferece um caminho para metodologias de triagem rápidas, de baixo custo e minimamente invasivas, com grande potencial para aplicações clínicas.

NÚCLEO DA ANÁLISE

Ideia de startup ou produto

Uma startup focada em 'Diagnósticos Espectrais Inteligentes', desenvolvendo plataformas de software e hardware que integram o algoritmo OGA para análise de dados espectroquímicos. O foco inicial seria em diagnósticos de doenças com alta demanda por triagem rápida e de baixo custo, como endometriose e doenças infecciosas, expandindo para controle de qualidade em alimentos e fármacos.

Aplicações práticas

Desenvolvimento de kits de diagnóstico avançados para doenças infecciosas (COVID-19, dengue) e condições crônicas (endometriose). Descoberta e validação de biomarcadores. Controle de qualidade e autenticação em diversas indústrias que demandam classificação de uma classe (ex: segurança alimentar, farmacêutica).

Potencial de mercado

Alto. O mercado global de diagnósticos in vitro é vasto e crescente, impulsionado pela demanda por testes mais rápidos, baratos e precisos. A tecnologia, por ser de baixo custo e minimamente invasiva, atende a uma necessidade crítica de mercado, especialmente em regiões com recursos limitados e para triagem em larga escala.

FUNDAMENTO CIENTÍFICO

Problema abordado

A necessidade de métodos de classificação robustos e precisos para dados espectroquímicos em diagnósticos clínicos, especialmente em cenários onde há limitação de dados de controle (classificação de uma classe). A otimização da seleção de variáveis é crucial para melhorar a performance diagnóstica e reduzir custos.

Metodologia

A pesquisa associa modelos de classificadores de classe única (DD-SIMCA e OCPLS com PRM) a uma nova estratégia de seleção de variáveis, o algoritmo genético de classe única (OGA). Essa combinação visa otimizar a extração de características de dados espectroquímicos para melhorar a acurácia na classificação de doenças.

Principais descobertas

O OGA melhora significativamente o desempenho da classificação quando associado a modelos OCC. O OGA-PRM-OCPLS alcançou 100% de sensibilidade para diagnóstico de COVID-19 e endometriose, enquanto o OGA-DD-SIMCA obteve 100% de sensibilidade para dengue. As variáveis selecionadas pelo OGA podem ser relacionadas a biomarcadores, com potencial para desenvolver metodologias de triagem rápidas, de baixo custo e minimamente invasivas.

LEITURA PARA GESTÃO PÚBLICA

A pesquisa pode subsidiar políticas públicas para a implementação de programas de triagem em massa para doenças como dengue e COVID-19, especialmente em áreas remotas ou com infraestrutura limitada, devido ao baixo custo e natureza minimamente invasiva da metodologia. Também pode guiar a regulamentação para validação e uso de ferramentas de diagnóstico baseadas em quimiometria e IA.

PARCERIA UNIVERSIDADE × EMPRESA

Parcerias entre a UECE e empresas de tecnologia médica (MedTech), farmacêuticas ou fabricantes de equipamentos espectrométricos para validação clínica em larga escala, otimização da integração do software OGA em dispositivos existentes e desenvolvimento de produtos diagnósticos comercializáveis. Colaboração com hospitais e laboratórios para testes e validação em ambiente real.

OPORTUNIDADES DERIVADAS

4 direções estratégicas identificadas

STARTUP · IMPACTO ALTO · HEALTHTECH

SpectroDx AI: Plataforma de Diagnóstico Rápido por Espectrometria

Desenvolver uma plataforma de diagnóstico baseada em IA que utiliza o algoritmo OGA para análise de dados espectroquímicos, oferecendo triagem rápida, de baixo custo e minimamente invasiva para doenças como COVID-19, dengue e endometriose. Foco inicial em mercados com alta demanda por diagnósticos acessíveis.

PARCERIA · IMPACTO ALTO · HEALTHTECH

Colaboração UECE-Indústria para Validação e Comercialização de Diagnósticos

Estabelecer parcerias entre a UECE e empresas de MedTech ou farmacêuticas para validar clinicamente o algoritmo OGA em larga escala, otimizar a integração com dispositivos espectrométricos existentes e desenvolver kits de diagnóstico comercializáveis, acelerando a chegada da tecnologia ao mercado.

POLÍTICA PÚBLICA · IMPACTO ALTO · GOVTECH

Programa Nacional de Triagem Inteligente para Saúde Pública

Propor e implementar políticas públicas que incentivem o uso de tecnologias de diagnóstico baseadas em IA e quimiometria (como o OGA) para programas de triagem em massa, visando a detecção precoce de doenças infecciosas e crônicas, reduzindo custos e melhorando o acesso à saúde em nível nacional.

PRODUTO CORPORATIVO · IMPACTO MÉDIO · SOFTWARE

Módulo de Análise OGA para Equipamentos Espectrométricos

Desenvolver um módulo de software ou hardware integrável para fabricantes de equipamentos espectrométricos, permitindo que seus dispositivos ofereçam análises de dados aprimoradas com seleção de variáveis otimizada pelo OGA, expandindo suas capacidades para aplicações de diagnóstico clínico e controle de qualidade industrial.

ABSTRACT ORIGINAL

One-class classifier (OCC) models are widely applied to solve classification problems where control or class modeling from a target class is necessary. In this study, OCC models such as Data Driven Soft Independent Modeling of Class Analogy (DD-SIMCA) and One-Class Partial Least Squares (OCPLS) were associated with a new variable selection strategy, the one-class genetic algorithm (OGA), for classification analyses in three clinical applications: COVID-19, endometriosis, and dengue samples. DD-SIMCA was implemented in a rigorous approach, using $\alpha = 0.05$, while OCPLS was performed with partial robust M-regression (PRM). For the three cases, a better classification performance was obtained using the OGA associated with the OCC model. The performance of the OGA-PRM-OCPLS showed better results for both COVID-19 and endometriosis cases when compared to DD-SIMCA, with a classification sensitivity of 100%. However, the best results for dengue classification were obtained by using the OGA-DD-SIMCA model (sensitivity = 100%). The selected variables obtained by the OGA can be used to relate this information to biomarkers capable of distinguishing between case and control groups. These findings have the potential to improve some disease diagnosis using chemometrics for the development of rapid, low-cost, and minimally invasive screening methodologies.

MATÉRIA PARA LEIGOS

Para leigos: Novo Algoritmo Otimiza o Diagnóstico de Doenças como COVID-19, Endometriose e Dengue

O cenário atual

Em diversas áreas, incluindo a saúde, é comum enfrentar desafios para classificar dados, especialmente quando temos muitas informações sobre um grupo (por exemplo, pessoas saudáveis) e poucas sobre o outro (pessoas com uma doença específica). Para resolver isso, são usados modelos chamados "classificadores de uma classe" (OCC, na sigla em inglês). Eles são importantes para criar modelos que identificam um grupo-alvo específico.

Existe uma busca constante por métodos de diagnóstico mais rápidos, baratos e que causem o mínimo de desconforto ao paciente. A análise de dados espectroquímicos, que envolve a interação da luz com a matéria para identificar substâncias, é uma ferramenta promissora nesse sentido.

O que os pesquisadores fizeram

Uma equipe de pesquisadores da Universidade Estadual do Ceará (UECE) desenvolveu uma nova estratégia para melhorar a análise de dados. Eles combinaram modelos de classificação de uma classe já existentes, como o DD-SIMCA e o OCPLS, com uma nova forma de selecionar variáveis, que chamaram de "algoritmo genético de uma classe" (OGA).

O objetivo foi aprimorar a capacidade de classificação em três aplicações clínicas importantes: amostras de pacientes com COVID-19, endometriose e dengue. Para o estudo, o modelo DD-SIMCA foi aplicado de forma rigorosa, enquanto o OCPLS foi executado com uma técnica específica chamada PRM (partial robust M-regression).

Como funciona na prática

A essência do trabalho é usar o algoritmo genético de uma classe (OGA) para identificar as variáveis mais importantes dentro dos dados espectroquímicos. Imagine que os dados têm muitas características, e o OGA atua como um "filtro inteligente" que seleciona apenas aquelas que são mais relevantes para distinguir entre, por exemplo, um paciente doente e um saudável.

Com essas variáveis selecionadas, os modelos de classificação (DD-SIMCA e OCPLS) conseguem trabalhar de forma mais eficiente e precisa. Essas variáveis destacadas pelo OGA podem, no futuro, ser associadas a biomarcadores, que são indicadores biológicos capazes de apontar a presença de uma doença.

Resultados e evidência

Os pesquisadores observaram que a combinação do algoritmo OGA com os modelos de classificação de uma classe (OCC) resultou em um desempenho de classificação superior em todas as três aplicações clínicas testadas. Isso significa que o sistema conseguiu identificar os casos com maior precisão.

Em particular, para os casos de COVID-19 e endometriose, a combinação OGA-PRM-OCPLS apresentou os melhores resultados, alcançando uma sensibilidade de 100% na classificação. A sensibilidade indica a capacidade do teste de identificar corretamente os casos positivos (ou seja, as pessoas que realmente têm a doença).

Para a classificação da dengue, o modelo OGA-DD-SIMCA foi o que obteve os melhores resultados, também com uma sensibilidade de 100%.

Implicações práticas

Essas descobertas têm um grande potencial para avançar o diagnóstico de doenças. Ao usar a quimiometria (a aplicação de métodos estatísticos a dados químicos) de forma mais eficaz, a pesquisa pode contribuir para o desenvolvimento de metodologias de triagem (rastreamento) que sejam rápidas, de baixo custo e minimamente invasivas. Isso poderia tornar o diagnóstico mais acessível e menos incômodo para os pacientes.

Limitações e próximos passos

O paper não detalha as limitações específicas do estudo ou os próximos passos da pesquisa. No entanto, o sucesso na seleção de variáveis e na alta sensibilidade sugere um caminho promissor para futuras validações e aplicações clínicas.

QUEM SÃO OS PESQUISADORES

Quem são os pesquisadores

Os autores deste estudo são José R. de Moraes Filho, Camilo de L. M. de Moraes, Anne B. F. Câmara e Kássio M. G. Lima. Todos são afiliados à Universidade Estadual do Ceará (UECE).

O paper não detalha as afiliações específicas de cada pesquisador dentro da universidade, seus vínculos prévios, sua trajetória acadêmica ou profissional, nem seus papéis individuais na condução deste estudo.

PALAVRAS-CHAVE

Classifier (UML) · Feature selection · Genetic algorithm · Pattern recognition (psychology) · Partial least squares regression · Class (philosophy) · Statistical classification · Sensitivity (control systems)