

Aprendizado de máquinas e Espectroscopia Raman para identificação de fungos

Resumo—A identificação da espécie de um fungo causador de doenças é determinante na escolha do tratamento mais adequado. Contudo tal tarefa pode ser desafiadora dada as semelhanças entre as espécies. Para tanto, pode ser utilizado o sequenciamento genético, entretanto, é uma técnica de elevado custo e requer equipamentos que não estão facilmente disponíveis. Este trabalho apresenta uma proposta de projeto para analisar dados do espectro Raman de algumas espécies de fungos através de técnicas de aprendizado de máquinas, de forma que, consiga-se identificar quais dados são relevantes para diferenciação e identificação de uma espécie específica dentre as analisadas.

I. INTRODUÇÃO

O Reino Fungi é composto pelas mais variadas formas de seres, desde micro-organismos até cogumelos. Contudo, quando deseja-se diferenciar espécies de mesmo gênero (táxon da classificação biológica) isto torna-se uma tarefa árdua, pois tais diferenças só são perceptíveis nas estruturas celulares destes seres, como pode ser verificado na figura 1.

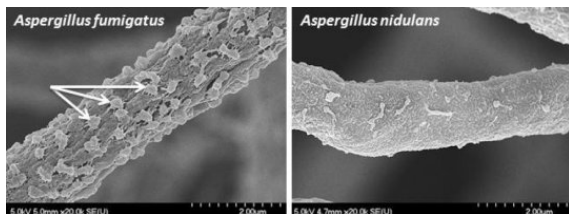


Figura 1. Diferença nas estruturas celulares das espécies fungos fumigatus e nidulans ambos do gênero Aspergillus. Fonte: [1]

Realizar o sequenciamento genético de uma amostra de células de um fungo é a forma mais assertiva de determinar a qual espécie o mesmo pertence. Mas, é um método com alto custo e requer um sequenciador genético a disposição. Por isto este trabalho propoe utilizar-se de dados de Espectroscopia Raman das células dos fungos para criar uma modelo de aprendizagem de máquina capaz de identificar a espécie de um fungo.

Nas próximas sessões deste trabalho serão detalhadas a significância do sinal da Espectroscopia Raman para o caso de estudo, técnicas de aprendizado de Máquina como PCA(Principal Component Analysis), Clusterização e Feature Selection e será proposta uma discussão do porquê estas técnicas podem trazer um valor significativo a materia de estudo, além de apresentar os resultados esperados ao fim desta pesquisa.

II. REVISÃO DA LITERATURA

Espectroscopia Raman é uma técnica que foi inspirada em um fenômeno observado experimentalmente por Chandrasekhara Venkata Raman. Tal fenômeno consiste no espalhamento da luz em diferentes frequências ao incidir um laser

sobre um material. Essas frequências luminosas se espalham devido a vibração entre as moléculas que foram excitadas pela luz incidida. Para cada molécula diferente uma intensidade luminosa pode ser mensurada, de modo que, a medida obtida revela a identidade da molécula. O conjunto dos valores mensurados, gera uma assinatura que é única (exemplo na figura 2, e por isso, capaz de revelar a estrutura molecular do material considerado.

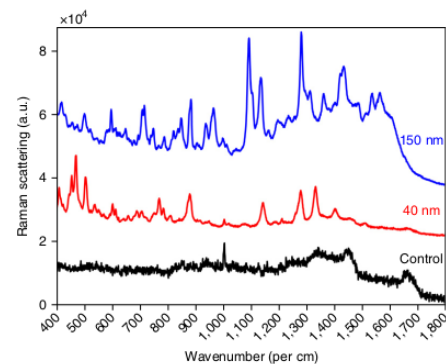


Figura 2. Exemplo de uma assinatura obtida através da Espectroscopia Raman. Fonte: [2]

Quando se realiza uma medição do espectro Raman, é preparada uma solução aquosa adicionando uma pequena quantidade do material que se deseja analisar e se adiciona um substrato a amostra (geralmente metais nobres) que por sua vez servirão como amplificadores da intensidade luminosa espalhada. A escolha do substrato é parte crucial na obtenção dos dados do espectro Raman conforme [3], pois ruídos podem ser gerados pelo substrato escolhido distorcendo os resultados.

Uma vez obtido os dados referentes ao espectro Raman é comum utilizar ferramental matemático para extrair da assinatura obtida ruídos advindos da fluorescência da composição material da amostra. Tal aspecto pode, por sua vez, mascarar a informação relevante do sinal que em geral é caracterizada por picos.

O PCA(Principal Component Analysis) é uma das ferramentas utilizadas para eliminar os ruídos do espectro Raman. Este método matemático é capaz de identificar em um espaço amostral de n dimensões quais destas carregam informação realmente relevantes para análise. Contudo, a eficiência deste método está no agrupamento de informação em poucas dimensões. Caso exista grande variação entre as componentes das amostras ao considerar apenas as de maior variância, muita informação relevante pode ser perdida comprometendo a análise que deseja-se realizar.

III. MATERIAIS E MÉTODOS

IV. DISCUSSÃO E RESULTADOS ESPERADOS

V. CONCLUSÃO

RECONHECIMENTO

Ao Professor Hugo Viera Neto, Ph.D. pelo excelente conteúdo ministrado na disciplina de Metodologia Científica do programa de pós graduação da Universidade Tecnológica Federal do Paraná.

REFERÊNCIAS

- [1] M. J. Lee, H. Liu, B. M. Barker, B. D. Snarr, F. N. Gravelat, Q. Al Abdallah, C. Gavino, S. R. Baistrocchi, H. Ostapska, T. Xiao, B. Ralph, N. V. Solis, M. Lehoux, S. D. Baptista, A. Thammahong, R. P. Cerone, S. G. Kaminskyj, M. C. Guiot, J. P. Latgé, T. Fontaine, D. C. Vinh, S. G. Filler, and D. C. Sheppard, "The Fungal Exopolysaccharide Galactosaminogalactan Mediates Virulence by Enhancing Resistance to Neutrophil Extracellular Traps," *PLoS Pathogens*, vol. 11, no. 10, pp. 1–22, 2015.
- [2] H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAinsh, N. Stone, and F. L. Martin, "Using Raman spectroscopy to characterize biological materials," *Nature Protocols*, 2016.
- [3] S. Seifert, V. Merk, and J. Kneipp, "Identification of aqueous pollen extracts using surface enhanced Raman scattering (SERS) and pattern recognition methods," *Journal of Biophotonics*, vol. 9, no. 1-2, pp. 181–189, 2016.