

Título do Projeto: Análise Comparativa do Desempenho de Ferramentas Computacionais para Perfilamento Automático de Metabólitos em Espectros de RMN ^1H

Pesquisador Responsável (Orientador): Dr. José Geraldo de Carvalho Pereira

1. Introdução

A metabolômica, o estudo dos metabólitos em sistemas biológicos, é crucial para entender processos fisiológicos, descobrir biomarcadores e desenvolver terapias. A Espectroscopia de Ressonância Magnética Nuclear (RMN) ^1H é uma técnica chave nesse campo devido à sua natureza quantitativa e não destrutiva. Contudo, a análise manual de espectros de RMN de amostras complexas é um processo lento, trabalhoso e subjetivo, representando um gargalo significativo. Para superar isso, foram desenvolvidas ferramentas computacionais automatizadas como BATMAN (Hao et al., 2012, 2014), Bayesil (Ravanbakhsh et al., 2015), NMR-QNet (Wang et al., 2023), SENNet (Xiao et al. 2024), nmRanalysis (Flores et al., 2025), que visam identificar e quantificar metabólitos. Apesar de promissoras, falta na literatura uma avaliação comparativa sistemática dessas ferramentas. Este projeto propõe realizar um *benchmark* padronizado dessas quatro ferramentas, utilizando dados sintéticos e experimentais, para avaliar seu desempenho e orientar a escolha da ferramenta mais adequada, identificando também áreas para desenvolvimento futuro.

2. Estado da Arte

Existem diversas ferramentas para automatizar o perfilamento de metabólitos via RMN. **BATMAN** usa uma abordagem Bayesiana, potente, mas computacionalmente intensiva. **Bayesil** é uma solução online automatizada focada em biofluidos específicos. **NMR-QNet** e **SENNet** aplicam Redes Neurais Convolucionais. A mais recente, **nmRanalysis**, é uma plataforma web integrativa com um sistema de recomendação baseado em aprendizado de máquina. Embora existam comparações pontuais (Wang et al., 2022; Gómez et al., 2014), falta um estudo comparativo direto e padronizado entre BATMAN, Bayesil, NMR-QNet, SENNet e nmRanalysis sob condições variadas e com métricas objetivas. Um *benchmark* rigoroso é essencial para guiar a comunidade científica e identificar necessidades de desenvolvimento.

3. Objetivos

- **Geral:** Avaliar e comparar o desempenho de BATMAN, Bayesil, NMR-QNet, SENNet e nmRanalysis na identificação e quantificação automática de metabólitos em espectros de RMN ^1H através de um *benchmark* padronizado.
- **Específicos:**
 - Revisar a literatura sobre ferramentas de perfilamento por RMN e metodologias de *benchmark*.
 - Preparar conjuntos de dados de RMN ^1H sintéticos (com *ground truth*) e experimentais (públicos/laboratório).
 - Definir métricas de desempenho para identificação (Acurácia, Precisão, Recall, F1, AUCs, etc.) e quantificação (RMSE, MAE, R^2 , etc.).
 - Instalar, configurar e aprender a operar as ferramentas computacionais.

- Executar as ferramentas nos datasets, registrando parâmetros e resultados.
- Analisar comparativamente os resultados, calculando métricas e identificando pontos fortes/fracos de cada ferramenta.
- Documentar e disseminar a metodologia e os resultados.

4. Metodologia:

O projeto seguirá as etapas:

- **Revisão Bibliográfica:** Análise da literatura sobre as ferramentas, metodologias de *benchmark* e RMN em metabolômica.
- **Preparação de Datasets:** Geração de dados sintéticos com variações controladas (nº metabólitos, sobreposição, S/R) e coleta/pré-processamento de dados experimentais (públicos e/ou do laboratório).
- **Definição de Métricas:** Seleção formal de métricas de classificação (identificação) e regressão (quantificação), além de métricas qualitativas/adicionais (tempo, usabilidade, robustez).
- **Operação das Ferramentas:** Instalação, aprendizado e execução sistemática das ferramentas nos datasets, com registro detalhado.
- **Análise Comparativa:** Processamento dos outputs, cálculo das métricas (via scripts Python), análise estatística e visualização dos resultados para comparação do desempenho.
- **Documentação e Disseminação:** Elaboração do relatório final, apresentações e discussão das implicações dos achados.

5. Cronograma de Atividades

- **Meses 1-2:** Revisão bibliográfica e definição detalhada dos datasets e métricas.
- **Meses 2-4:** Preparação e curadoria dos datasets (sintéticos e experimentais), definição final das métricas.
- **Meses 4-5:** Instalação, configuração e aprendizado das ferramentas (BATMAN, Bayesil, NMR-QNet, SENNet, nmRanalysis).
- **Meses 5-9:** Execução das ferramentas nos datasets preparados e coleta dos resultados.
- **Meses 5-6:** Elaboração do relatório parcial
- **Meses 9-11:** Processamento dos resultados, cálculo das métricas e análise comparativa.
- **Meses 11-12:** Documentação, redação do relatório final e preparação de materiais de disseminação.

6. Referências:

Flores, J. E., Prymolenna, A. V., Lewis, L. A., et al. (2025). nmRanalysis: An Open-Source Web Application for Semi-automated NMR Metabolite Profiling. *Analytical Chemistry*.

Gómez, J., Brezmes, J., Mallol, R., et al. (2014). rDolphin: open software for robust quantification of metabolites in complex ^1H -NMR spectra. *Analytical and Bioanalytical Chemistry*, 406(30), 7967–7976.

Hao, J., Astle, W., De Iorio, M., & Ebbels, T. M. D. (2012). BATMAN—an R package for Bayesian automated metabolite analysis for NMR spectra. *Bioinformatics*, 28(15), 2088–2090.

Hao, J., Liebeke, M., Astle, W., et al. (2014). Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra. *Nature Protocols*, 9(6), 1416–1427.

Ravanbakhsh, S., Liu, P., Bjorndahl, T. C., et al. (2015). Accurate, Automated NMR Metabolite Identification and Quantification Without Reference Spectra. *PLOS Computational Biology*, 11(5), e1004006.

Wang, W., Ma, L. H., Maletic-Savatic, M., & Liu, Z. (2023). NMRQNet: a deep learning approach for automatic identification and quantification of metabolites using Nuclear Magnetic Resonance (NMR) in human plasma samples. *bioRxiv*.

Wang, X., Mickiewicz, B., Thompson, G. C., et al. (2022). Performance evaluation of manual profiling, BATMAN and rDolphin for quantitative NMR metabolomics. *Metabolites*, 12(3), 227.

Xiao, X., Wang, Q., Chai, X. et al. (2024). Using neural networks to obtain NMR spectra of both small and macromolecules from blood samples in a single experiment. *Commun Chem* 7, 167.