

# REVISÃO BIBLIOMÉTRICA DA ESPECTROSCOPIA NO INFRAVERMELHO MÉDIO (MIR) ASSOCIADA AO APRENDIZADO DE MÁQUINA NA TRIAGEM DA ESCLEROSE MÚLTIPLA

## *BIBLIOMETRIC REVIEW OF MID-INFRARED (MIR) SPECTROSCOPY ASSOCIATED WITH MACHINE LEARNING IN MULTIPLE SCLEROSIS SCREENING*

Anne Louise S. Torres<sup>1,2</sup>; Raí dos Santos Santiago<sup>2</sup>; Márcia Helena Cassago do Nascimento<sup>2</sup>; Livia Carla de Melo Rodrigues<sup>2</sup>; Valerio Garrone Barauna<sup>2</sup>

<sup>1</sup>Instituto Federal do Espírito Santo (IFES), Espírito Santo, Brasil.

<sup>2</sup>Universidade Federal do Espírito Santo (UFES), Espírito Santo, Brasil.

### Resumo

**Introdução:** O campo de estudo denominado bioespectroscopia emprega técnicas de espectroscopia vibracional na região do infravermelho médio (MIR) para o estudo de problemas biológicos. Nesse contexto, a espectroscopia MIR tem sido proposta como uma nova ferramenta para triagem de doenças com diagnósticos complexos, como a esclerose múltipla (EM), por meio da análise de biofluidos e métodos de aprendizado de máquina. Na literatura, há artigos científicos que construíram modelos de detecção de EM por meio dessa técnica com diferentes abordagens. A EM é uma doença crônica neurodegenerativa e autoimune que acomete principalmente o sistema nervoso central. **Objetivo:** Realizar uma análise bibliométrica de artigos sobre triagem para EM por espectroscopia MIR. **Metodologia:** Foram utilizados artigos em inglês que construíram modelos de diagnóstico para EM por meio da espectroscopia MIR. A busca foi realizada com as palavras-chave “multiple sclerosis” AND “infrared spectroscop\*” nas bases de dados Pubmed, Web of Science e Scopus, desde o período de indexação do primeiro artigo até maio de 2025. **Resultados:** Foram incluídos 6 artigos para compor a presente revisão. Os trabalhos incluídos utilizaram diferentes metodologias para construção de modelos de classificação para EM. O biofluido mais utilizado foi o sangue e a região espectral mais utilizada foi a de 1500-600 cm<sup>-1</sup>, denominada de *fingerprint*. Além disso, foi possível observar que o grupo amostral utilizado nos trabalhos é pequeno. **Conclusão:** A espectroscopia MIR associada à análise multivariada e aprendizado de máquina mostrou-se como uma ferramenta promissora para triagem da esclerose múltipla. No entanto, a falta de padronização metodológica limita sua aplicação clínica.

**Palavras-chave:** Bioespectroscopia, Biofluidos, Aprendizado de máquina, Esclerose Múltipla, Bibliometria.

## *Abstract*

**Introduction:** The field of study known as biospectroscopy employs vibrational spectroscopy techniques in the mid-infrared (MIR) region to investigate biological problems. In this context, MIR spectroscopy has been proposed as a new tool for screening diseases with complex diagnoses, such as multiple sclerosis (MS), through the analysis of biofluids and machine learning methods. In the literature, there are scientific articles that have developed MS detection models using this technique with different approaches. MS is a chronic neurodegenerative and autoimmune disease that primarily affects the central nervous system. **Objective:** To conduct a bibliometric analysis of articles on MS screening using MIR spectroscopy. **Methodology:** English-language articles that developed diagnostic models for MS using MIR spectroscopy were included. The search was carried out using the keywords “multiple sclerosis” AND “infrared spectroscopy\*” in the PubMed, Web of Science, and Scopus databases, from the indexing of the first article up to May 2025. **Results:** Six articles were included in this review. The studies employed different methodologies to construct classification models for MS. The most commonly used biofluid was blood, and the most frequently analyzed spectral region was 1500–600 cm<sup>-1</sup>, known as the fingerprint region. Moreover, it was observed that the sample sizes used in these studies were small. **Conclusion:** MIR spectroscopy combined with multivariate analysis and machine learning has proven to be a promising tool for multiple sclerosis screening. However, the lack of methodological standardization limits its clinical application.

**Keywords:** Biospectroscopy, Biofluids, Machine Learning, Multiple Sclerosis, Bibliometrics

Recebido em: 28-07-2025

Publicado em: 08-10-2026

## *Autor correspondente*

Anne Louise Silva Torres

Endereço: Avenida Ministro Salgado Filho nº 1.000, Soteco, CEP: 29106-010, Vila Velha, ES, Brasil.

Email: [annelouise1106@gmail.com](mailto:annelouise1106@gmail.com)

## 1. Introdução

O campo da bioespectroscopia emprega

técnicas de espectroscopia vibracional na região do infravermelho para o estudo de problemas de base biológica. A

espectroscopia na região do infravermelho (IR) baseia-se no estudo da absorção, emissão, dispersão e espalhamento de luz da radiação emitida na região do infravermelho com a matéria em diferentes comprimentos de onda<sup>1</sup>. Assim, por meio de uma análise pela espectroscopia IR é possível obter informações sobre a estrutura, composição e dinâmica das moléculas presentes em uma substância. Dentre as técnicas de IR aplicadas às ciências da saúde destacam-se a espectroscopia na região do infravermelho médio (MIR)<sup>2</sup>, espectroscopia na região do infravermelho próximo (NIR)<sup>3</sup> e espectroscopia Raman.<sup>4</sup>

A espectroscopia MIR permite a análise simultânea de diferentes moléculas em matrizes biológicas complexas, fornecendo uma impressão digital bioquímica global da amostra. Essa característica torna a técnica especialmente útil para triagem clínica, pois dispensa etapas de separação e reagentes específicos, é capaz de detectar alterações em proteínas, lipídios, ácidos nucleicos e carboidratos em biofluidos como plasma, soro, urina e saliva.<sup>5,6,7,8</sup> Dessa forma, a espectroscopia MIR tem sido considerada uma técnica multiômica promissora para a triagem não invasiva de doenças.

Diante da complexidade do diagnóstico e da ausência de um biomarcador específico para a esclerose múltipla (EM)<sup>9,10,11</sup> a espectroscopia MIR, associada a métodos de análise multivariada e reconhecimento de padrões tem sido proposta na literatura como um possível método de triagem clínica. A EM é uma doença crônica neurodegenerativa, autoimune sistêmica, caracterizada por inflamação com desmielinização e proliferação astrogliar que acomete principalmente o sistema nervoso

central (SNC).<sup>13</sup> O diagnóstico atual baseia-se nos critérios de McDonald<sup>10</sup> e na integração das observações clínicas dos sintomas, ressonância magnética por imagem do SNC e análise laboratorial do líquido cefalorraquidiano.<sup>11</sup> Entretanto, esses métodos são caros, invasivos e nem sempre acessíveis, o que reforça a necessidade de estratégias alternativas e menos invasivas para a triagem e o acompanhamento da doença.

Embora diversos estudos relatem seu potencial diagnóstico utilizando diferentes biofluidos e abordagens espectrais, a literatura carece de uma revisão sistemática comparando esses trabalhos. Assim, faz-se necessária uma análise aprofundada e detalhada desses estudos para que sejam comparados e que se saibam as fraquezas e os pontos fortes das metodologias utilizadas com o intuito de aproximar o uso da espectroscopia MIR na prática clínica.

Considerando a expansão acelerada das publicações científicas e a fragmentação dos fluxos de pesquisa<sup>12</sup>, as revisões bibliométricas tornam-se fundamentais para mapear o desenvolvimento científico de um tema. Essa abordagem oferece uma análise quantitativa e reprodutível da literatura, permitindo identificar tendências, autores e instituições de destaque.<sup>13,14,15</sup> Assim, o presente estudo teve como objetivo mapear, por meio de uma análise bibliométrica baseada nas ferramentas Bibliometix<sup>16</sup> e Methodi Ordinatio<sup>17</sup>, os artigos publicados sobre o uso da espectroscopia MIR na triagem da esclerose múltipla, destacando as matrizes biológicas analisadas, as abordagens espectrais e as técnicas de modelagem empregadas.

## 2. Metodologia

### 2.1 BASES DE DADOS E PALAVRAS-CHAVES

Para a realização deste trabalho foi realizada uma análise bibliométrica pelo método Bibliometrix<sup>16</sup> (versão 4.1.2) implementado pelo *software* RStudio. A busca considerou a combinação das palavras-chaves “*multiple sclerosis*” AND “*infrared spectroscopy*” nos bancos de dados *Pubmed*, *Web of science* e *Scopus*. Foram incluídos artigos publicados desde o período de indexação do primeiro artigo até maio de 2025.

Os critérios de inclusão adotados foram: artigos originais publicados em inglês, estudos que utilizassem a técnica MIR, estudos que utilizassem biofluidos ou tecidos biológicos humanos e possuíssem uma abordagem de triagem para esclerose múltipla. Como critérios de exclusão foram adotados: revisões de literatura, resumos de conferências, capítulos de livros, cartas ao editor e trabalhos não relacionados à aplicação clínica ou triagem da EM.

### 2.2 PROCESSAMENTO E ORGANIZAÇÃO DOS DADOS

Os artigos encontrados em cada base de dados foram extraídos em formato Bibtex (.bib), nas bases de dados Web of Science (WoS) e Scopus, e no formato texto (.txt) na base de dados Pubmed. Posteriormente, foram importados para o *software* RStudio onde os dados foram organizados e retirados os artigos duplicados. Em seguida, os dados foram estruturados e organizados para a análise bibliométrica. O pacote bibliometrix<sup>16</sup> foi utilizado para visualizar as revistas que publicam esse tema,

mapa referente ao número de produções científicas por países, número de publicações ao longo do ano e nuvem de palavras mais citadas nos artigos que abrangem as palavras-chave selecionadas.

Após os processamentos realizados no R Studio, os artigos foram inspecionados individualmente por uma planilha no *software* Microsoft Excel (*Science Mapping*). A priorização dos artigos mais relevantes foi realizada de acordo com o método *Methodi Ordinatio*<sup>17</sup> (equação 1). O índice de *Ordinatio* avalia o grau de relevância dos artigos presentes na literatura por meio da análise da quantidade de citações, ano de publicação do artigo em relação ao ano de realização da análise bibliométrica e fator de impacto do periódico em que o artigo foi publicado.

$$InOrdinatio = x \cdot FI + b \cdot [10 - (\text{Ano da Pesquisa} - \text{Ano da publicação})] + y \cdot Ci \quad (1)$$

Em que:

FI: Fator de Impacto no JCR do periódico onde o artigo foi publicado.

Ano da pesquisa: foi utilizado para esse estudo o ano de 2025.

Ci: Quantidade de Citações do artigo.

x, b e y: pesos atribuídos a cada parâmetro (neste estudo, x=1, b=1 e y=1).

Os artigos foram ordenados de forma decrescente pelo valor de *InOrdinatio* e foram submetidos à leitura integral para confirmar que o estudo se tratava do uso da espectroscopia MIR para a triagem clínica de EM.

### 2.3 EXTRAÇÃO E SÍNTESE DAS INFORMAÇÕES

Para os artigos incluídos na revisão, foi realizado o fichamento detalhado em planilha Excel para extrair as informações dos artigos, como: título, autores, ano de publicação, matriz biológica analisada, grupo amostral, equipamento de espectroscopia MIR utilizado, configurações do equipamento, aquisição do espectro, região espectral analisada, pré-processamentos utilizados, métodos de inteligência artificial, análises estatísticas, softwares utilizados, áreas relatadas como importantes para discriminação dos grupos com EM, parâmetros de avaliação dos modelos e conclusões relevantes de cada artigo.

## 2.4 FLUXO DE SELEÇÃO (PRISMA)

A seleção dos estudos foi descrita conforme o protocolo PRISMA 2020<sup>18,19</sup>, incluindo a etapa de identificação: número total de registros obtidos nas três bases de dados; etapa de triagem: exclusão de duplicatas e aplicação dos critérios de exclusão de acordo com o tipo do artigo; etapa de elegibilidade: aplicação dos critérios de exclusão de acordo com a leitura do título e resumo; etapa de inclusão (número final de artigos incorporados à análise bibliométrica). O fluxograma PRISMA (Figura 1) ilustra as etapas e resultados obtidos em cada fase.

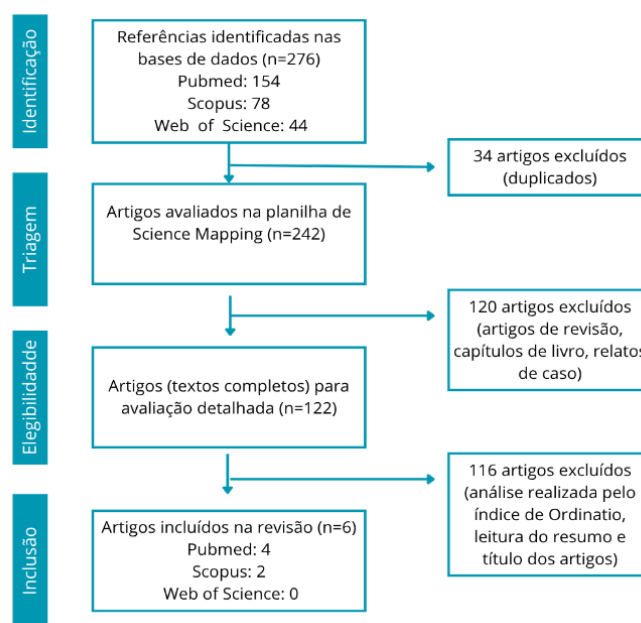


FIGURA 1 – Fluxograma da seleção dos artigos incluídos na revisão. Fonte: Os autores.

## 2.5 AVALIAÇÃO DA QUALIDADE DOS ESTUDOS

A qualidade dos estudos incluídos foi avaliada de forma adaptada a partir do instrumento de Avaliação da Qualidade de Estudos de Precisão Diagnóstica (QUADAS-2, do inglês

*Quality Assessment of Diagnostic Accuracy Studies*).<sup>20</sup> Foram avaliados cinco domínios: seleção dos participantes, aquisição dos espectros, pré-processamentos espectrais, modelagens estatísticas realizadas e clareza na apresentação dos resultados.

Observou-se que a seleção dos participantes para realização dos estudos foi apresentada de forma clara quanto à definição dos grupos controle e EM. No entanto, 2 estudos não detalharam os critérios de exclusão adotados. A aquisição dos espectros foi satisfatoriamente descrita em todos os artigos incluídos nesta revisão, descritas as regiões espectrais utilizadas e o tipo de espectrômetro. Entretanto, alguns artigos não apresentaram informações sobre os parâmetros instrumentais adotados para aquisição espectral. Os pré-processamentos aplicados aos espectros foram descritos em todos os estudos, mas nem sempre descritos de forma completa para garantir a reprodutibilidade. Quanto à modelagem estatística, todos os estudos aplicaram métodos multivariados ou de aprendizado de máquina para discriminação dos grupos controle vs. EM, embora faltassem informações sobre os métodos de validação cruzada adotados na construção dos modelos, a metodologia utilizada para validação dos modelos e o controle de *overfitting*. A clareza na apresentação dos resultados foi considerada adequada, uma vez que os artigos apresentaram gráficos e tabelas bem estruturados, mas alguns trabalhos carecem de discussão crítica quanto às limitações e incertezas dos modelos preditivos construídos. Dessa forma, os estudos apresentaram baixo risco de viés, com limitações pontuais

relacionadas à clareza da apresentação das metodologias aplicadas, sem comprometer a validade global das conclusões.

### 3. Resultados

O método de bibliometria permitiu, pela interface do Bibliometrix, analisar de forma sistemática e gráfica os artigos encontrados. Dessa forma, foi possível identificar as revistas nas quais esses estudos foram publicados, quais sejam: *Scientific Reports* com duas publicações, *Sensors*, *International Journal of Molecular Sciences*, *Biochemical and Biophysical Research Communications* e *Biochimica et Biophysica Acta*, com 1 publicação cada.

Além disso, foi possível identificar os países com mais produções científicas acerca do tema. Identificou-se, de acordo com a localização do autor correspondente, que as publicações dentro do tema proposto estão concentradas na Europa e na América do Norte: Itália, Polônia, Turquia, Eslováquia, França e Canadá. No mapa, é possível identificar que todos os países em destaque apresentam a mesma intensidade de cor de preenchimento, indicando que os países possuem o mesmo número de publicações no tema abordado, cada país com 1 documento indexado (Fig.2).





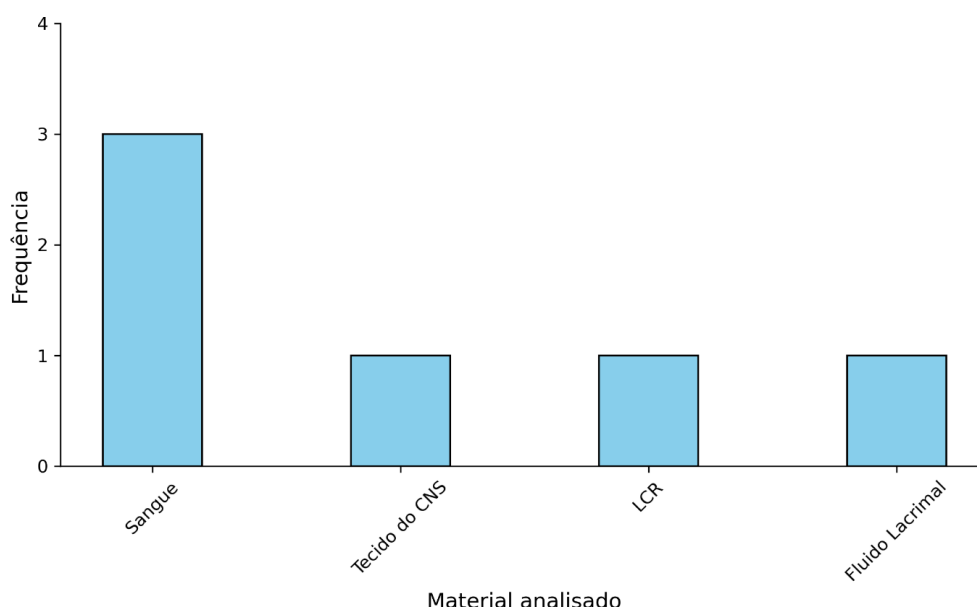
|                                                                                                                                                                  |      |    |                                                                         |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|----|-------------------------------------------------------------------------|----|
| <i>Rapid Discrimination of Neuromyelitis Optica Spectrum Disorder and Multiple Sclerosis Using Machine Learning on Infrared Spectra of Sera</i>                  | 2022 | 11 | <i>International Journal of Molecular Sciences</i>                      | A2 |
| <i>Experimental Analysis of Tear Fluid and Its Processing for the Diagnosis of Multiple Sclerosis</i>                                                            | 2023 | 13 | <i>Sensors</i>                                                          | A2 |
| <i>ATR-FTIR spectroscopy of plasma supported by multivariate analysis discriminates multiple sclerosis disease</i>                                               | 2023 | 10 | <i>Scientific Reports</i>                                               | A1 |
| <i>Relapsing-Remitting Multiple Sclerosis diagnosis from cerebrospinal fluids via Fourier transform infrared spectroscopy coupled with multivariate analysis</i> | 2018 | 72 | <i>Scientific Reports</i>                                               | A1 |
| <i>Infrared spectroscopic characterisation of multiple sclerosis plaques in the human central nervous system</i>                                                 | 1993 | 64 | <i>Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease</i> | A2 |
| <i>Spectral signature of multiple sclerosis. Preliminary studies of blood fraction by ATR FTIR technique</i>                                                     | 2022 | 17 | <i>Biochemical and Biophysical Research Communications</i>              | A2 |

A partir da leitura e análise sistemática dos artigos incluídos nessa revisão bibliométrica foi possível identificar: matriz biológica mais utilizada; média amostral por técnica; região espectral utilizada em cada metodologia; técnicas utilizadas para o tratamento dos dados; métodos de aprendizagem de máquina utilizados; métricas de avaliação dos modelos de classificação desenvolvidos.

A figura 4 representa graficamente a utilização da matriz biológica para cada metodologia descrita nos artigos selecionados. Consta-se que a matriz biológica mais explorada foram as porções do sangue (plasma ou soro) com 3 artigos, destes, 1 artigo analisou as duas frações sanguíneas simultaneamente.<sup>21</sup> Conforme os autores dos artigos, a escolha desse biofluido é justificada por:

ser minimamente invasivo quando comparado ao biofluido padrão-ouro para o diagnóstico de EM que é o líquido cefalorraquidiano; e pela sua composição que consegue refletir as condições de diferentes órgãos e tecidos corporais

simultaneamente. Além disso, 1 artigo utilizou o próprio líquido cefalorraquidiano, 1 artigo utilizou fluido lacrimal e 1 artigo utilizou o tecido do SNC como matriz biológica.



**FIGURA 4** - Distribuição de artigos por matriz biológica. Fonte: Os autores.

Kołodziej *et al.*<sup>21</sup> investigaram se havia diferença entre usar plasma ou soro para classificar indivíduos saudáveis e com EM. Para isso, aplicaram Análise de Cluster Hierárquico e Análise de Componentes Principais a espectros de infravermelho médio. A região de *biofingerprint* (1700–900  $\text{cm}^{-1}$ ) mostrou uma clara distinção entre plasma e soro, mais significativa do que entre as condições clínicas (saudável vs. EM). A principal diferença foi em 1652  $\text{cm}^{-1}$ , associada a proteínas  $\alpha$ -helicoidais, especialmente a albumina, predominante no soro e rica nesse tipo de estrutura.

Os artigos participantes apresentados nesta revisão relataram uma média de 89 participantes, variando de 12 a 235 participantes. Na Tabela 2, está apresentada a quantidade de indivíduos

incluídos em cada estudo de acordo com o grupo amostral. Yonar *et al.*<sup>22</sup> utilizou para compor o grupo doente duas formas clínicas diferente da EM, para isso utilizou 35 pacientes com Esclerose Múltipla Recorrente-Remitente (EMRR) e 30 pacientes com Síndrome Clinicamente Isolada (CIS) que ainda não possuem o diagnóstico da EM, mas estão altamente propensos a desenvolver a doença, o que pode contribuir para um diagnóstico precoce. Já, El Khoury *et al.*<sup>23</sup> em seu estudo construíram modelos de classificação para EM e para o Transtorno do Espectro da Neuromielite Óptica comparada com o grupo controle e entre elas. Isso devido a ambas serem doenças inflamatórias e desmielinizantes autoimunes do sistema nervoso central o que dificulta a distinção dentre elas, impossibilitando o diagnóstico clínico.

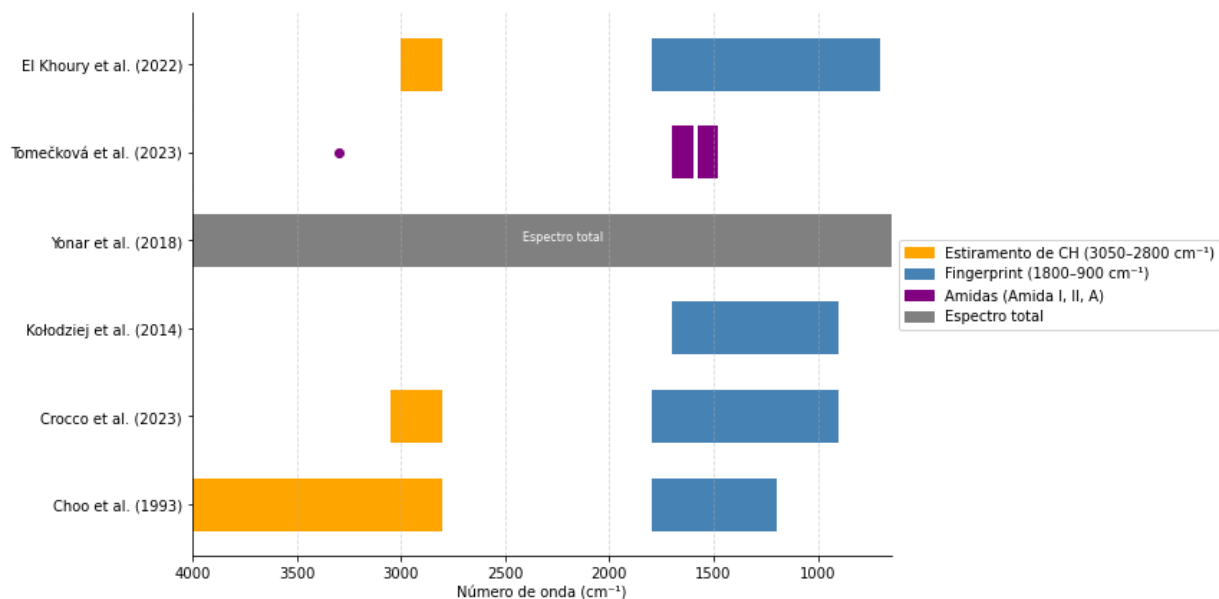
TABELA 2 – Número amostral utilizado nos artigos incluídos na revisão.

| Artigo científico                            | Número amostral                                         |
|----------------------------------------------|---------------------------------------------------------|
| Crocco <i>et al.</i> (2023) <sup>24</sup>    | n=85. EM= 45 e CT=40.                                   |
| Kołodziej <i>et al.</i> (2022) <sup>21</sup> | n=22. EM=11 e CT=11.                                    |
| Yonar <i>et al.</i> (2018) <sup>22</sup>     | n=89. EM=65 e CT=24.                                    |
| Tomecková <i>et al.</i> (2023) <sup>25</sup> | n=12. EM=6 e CT=6.                                      |
| El Khoury <i>et al.</i> (2022) <sup>23</sup> | n= 235. EM=60; CT= 80;<br>NMOSD=60.                     |
| Choo, L. <i>et al.</i> (1993) <sup>26</sup>  | <i>National Neurological<br/>Research Specimen Bank</i> |

Nota: Esclerose Múltipla (EM), Controle (CT), Transtorno do Espectro da Neuromielite Óptica (NMOSD, do inglês *Neuromyelitis Optica Spectrum Disorder*).

Ao analisar a região espectral utilizada por cada artigo foi possível observar que não houve consenso entre os estudos, como observado na figura 5. Crocco *et al.*<sup>24</sup> utilizaram a região de alto número de onda (3050-2800 cm<sup>-1</sup>) e *fingerprint* (1800-900 cm<sup>-1</sup>). Kołodziej *et al.*<sup>21</sup> utilizaram apenas a região de *fingerprint* (1700-900 cm<sup>-1</sup>) enquanto Yonar *et al.*<sup>22</sup> optaram por utilizar toda região espectral (4000–650 cm<sup>-1</sup>). Já Tomecková *et al.*<sup>25</sup> escolheram uma região mais

específica, a região relacionada as proteínas: Amida I (1600–1700 cm<sup>-1</sup>), Amida II (1480-1575cm<sup>-1</sup>) e Amida A (3300cm<sup>-1</sup>), que estão presentes tanto na região de *fingerprint* quanto na região de alto número de onda. El Khoury *et al.*<sup>23</sup> utilizaram a região de 3000–2800 cm<sup>-1</sup> e 1800–700 cm<sup>-1</sup> que compreendem quase ao espectro total. Por fim, Choo *et al.*<sup>26</sup> utilizaram a região espectral de 1200–1800cm<sup>-1</sup> e 2800–4000 cm<sup>-1</sup>.



**FIGURA 5** – Regiões espectrais utilizadas em cada estudo incluídos nesta revisão. Fonte: Os autores.

Além disso, foram avaliadas as metodologias de análise dos espectros utilizadas pelos autores dos artigos científicos incluídos nessa revisão para construção do modelo de classificação.

Para isso, considerou-se o pré-processamento utilizado pelos autores, o método de aprendizagem de máquina aplicado e a métrica de classificação obtida em cada estudo (Tabela 3).

**TABELA 3** – Comparação das metodologias de acordo com o modelo construído e as métricas alcançadas.

| Artigo científico                            | Pré-processamento                                      | Modelo               | Métricas                                                                                                                               |
|----------------------------------------------|--------------------------------------------------------|----------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| Crocco <i>et al.</i> (2023) <sup>24</sup>    | Correção de linha de base, normalização, 2ª Derivada   | PCA-LDA, PLS-DA e RF | Fingerprint (RF): 78% sensibilidade e 83% de especificidade. Alto número de ondas (PLS-DA): 67% sensibilidade e 83% de especificidade. |
| Kołodziej <i>et al.</i> (2022) <sup>21</sup> | Normalização min-max na região de amida I, 2ª Derivada | HCA e PCA            | Plasma: PC1 28% e PC2 25%. Soro: PC1 56% e PC2 17%.                                                                                    |

|                                              |                                                                                       |                |                                                                                                                        |
|----------------------------------------------|---------------------------------------------------------------------------------------|----------------|------------------------------------------------------------------------------------------------------------------------|
| Yonar <i>et al.</i> (2018) <sup>22</sup>     | Correção de linha de base e normalização em relação à banda de amida A                | PCA e HCA      | PC1 86% e PC2 7%.<br>AUC 0,86                                                                                          |
| Tomecková <i>et al.</i> (2023) <sup>25</sup> | -                                                                                     | Análise visual | Apenas comparação pela localização (número de onda) das bandas selecionadas dos espectros médios dos grupos (EM e CT). |
| El Khoury <i>et al.</i> (2022) <sup>23</sup> | Correção de linha de base, normalização e 2ª Derivada                                 | RF             | Grupos controle, grupo NMSDO e grupo EM com 100% para curva ROC, sensibilidade, especificidade e precisão.             |
| Choo <i>et al.</i> (1993) <sup>26</sup>      | Subtração interativa do espectro de vapor de água e diamante; deconvolução de Fourier | Análise visual | Apenas comparação pela localização (número de onda) das bandas dos espectros dos grupos (EM e CT).                     |

Nota: Análise de Componentes Principais (PCA, do inglês *Principal Components Analysis*), Análise Discriminante Linear (LDA, do inglês *Linear Discriminante Analysis*), Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA, do inglês *Partial Least Squares Discriminant Analysis*), Floresta Aleatória (RF, do inglês *Random Forest*).

#### 4. Discussão

Crocco *et al.*<sup>24</sup> conseguiram em seu modelo de classificação com Random Forest 78% de sensibilidade e 83% de especificidade utilizando a região de fingerprint e com o modelo de PLS-DA 67% sensibilidade e 83% especificidade na região de alto número de onda. Os autores destacaram que a região de 1750-1725 cm<sup>-1</sup>, atribuída à banda de estiramento éster C=O dos lipídios, é

mais intensa no grupo EM enquanto o pico da banda de amida I apresentou intensidade reduzida. Além disso, os autores discutem que o aumento na intensidade da banda C=O em relação à intensidade do pico de absorção de CH<sub>2</sub> em 1468 cm<sup>-1</sup> observado em seu estudo poderia estar associado à presença de lipídios oxidados produzidos por radicais, que supostamente desempenham um papel na patogênese da esclerose

múltipla. Como conclusão, relataram que existe um aumento na relação lipídio/proteína nos pacientes com EM em comparação ao grupo controle.

Kołodziej *et al.*<sup>21</sup> com PCA conseguiram explicar 53% da variância dos dados com o plasma na região de 1300- 900  $\text{cm}^{-1}$  e com o soro 73% ao utilizar a região de 1700-900  $\text{cm}^{-1}$ . Na análise que utilizou o plasma a melhor região para diferenciação, segundo os autores, foi entre 1360-900  $\text{cm}^{-1}$  já para o soro foi observado uma alteração na posição da banda de amida III (1265 $\text{cm}^{-1}$  CT vs. 1284 $\text{cm}^{-1}$  EM). Os autores relataram que a análise supervisionada revelou padrões de identificação claros no plasma e nas amostras de soro, indicando que a amida III, folha b e estruturas da amida I, respectivamente, são as mais proeminentes na diferenciação de EM.

Yonar *et al.*<sup>22</sup> também realizaram uma análise exploratória dos dados com PCA na região de 3025–2800  $\text{cm}^{-1}$  e 813–775  $\text{cm}^{-1}$  e conseguiram explicar 93% da variância dos dados e na região de 1950–2830  $\text{cm}^{-1}$  que explicou 73% da variância dos dados. O modelo desenvolvido pelos autores com o SIMCA obteve uma área abaixo da curva de 0,86 para a classificação das amostras. Os autores relataram em seu trabalho que as diferenças biomoleculares, como a peroxidação dos lipídios de membrana, dano oxidativo de proteínas e aumento do conteúdo de ácidos nucleicos são relevantes para o diagnóstico e permitem a discriminação das amostras doentes pelo LCR. Segundo os autores, há uma diminuição significativa das proporções de amida I/amida II e lipídio/proteína para o grupo com EM. Por fim, os autores relataram uma banda exclusiva em espectros de pacientes com EM na região de 795  $\text{cm}^{-1}$ , que de acordo com a literatura pertence à conformação

C3'-endo/syn da guanina no Z-.

O estudo de Tomecková *et al.*<sup>25</sup> realizou apenas uma comparação espectral entre bandas selecionadas pelos autores, para isso utilizaram os espectros médio dos grupos EM vs. Controle. Como resultado, obtiveram diferenças na posição das bandas de Amida A (Controle: 3297  $\text{cm}^{-1}$ ; EM: 3297  $\text{cm}^{-1}$ ), Amida I (Controle: 1658  $\text{cm}^{-1}$ ; EM: 1657  $\text{cm}^{-1}$ ) e Amida II (Controle: 1545  $\text{cm}^{-1}$ ; EM: 1543  $\text{cm}^{-1}$ ). Como conclusão, os autores destacaram que os pacientes com EM apresentaram proteomas lacrimais alterados, que segundo eles pode ser utilizado como um indicador de distúrbios relacionados à EM e inflamação do SNC.

El Khoury *et al.*<sup>23</sup> alcançaram para os grupos controle, NMSDO e EM 100% para curva ROC, sensibilidade, especificidade e precisão com o modelo de Random Forest (RF). O modelo desenvolvido pelos autores utilizou 20 principais regiões espectrais para discriminação dos grupos: 2954  $\text{cm}^{-1}$ , 2881  $\text{cm}^{-1}$ , 1535  $\text{cm}^{-1}$ , 1698  $\text{cm}^{-1}$ , 1695  $\text{cm}^{-1}$ , 2952  $\text{cm}^{-1}$ , 1565  $\text{cm}^{-1}$ , 1403  $\text{cm}^{-1}$ , 2938  $\text{cm}^{-1}$ , 1619  $\text{cm}^{-1}$ , 952  $\text{cm}^{-1}$ , 2917  $\text{cm}^{-1}$ , 2933  $\text{cm}^{-1}$ , 2975  $\text{cm}^{-1}$ , 1689  $\text{cm}^{-1}$ , 2877  $\text{cm}^{-1}$ , 2969  $\text{cm}^{-1}$ , 1614  $\text{cm}^{-1}$ , 2875  $\text{cm}^{-1}$ , 2867  $\text{cm}^{-1}$ . Foi possível observar que a grande maioria dessas regiões se encontra na faixa espectral entre 2.977-2.881  $\text{cm}^{-1}$ , dominada pela absorção da vibração de estiramento antissimétrico dos grupos  $\text{CH}_3$  e  $\text{CH}_2$  segundo os autores. Em resumo, o artigo sugeriu que existe uma grande diferença na composição de lipídios e/ou lipoproteínas, e estas são responsáveis pela distinção das amostras.

Por fim, o último artigo incluído nessa revisão de Choo *et al.*<sup>26</sup> foi um estudo pioneiro na utilização da espectroscopia MIR para triagem de EM, utilizando com matriz biológica tecido do SNC. Os

autores observaram uma discriminação espectral entre substância branca, cinzenta e placas crônicas de EM, além de alterações em bandas espectrais relacionadas a lipídios, proteínas e DNA quando comparados espectros de pacientes controle vs. EM.

De forma geral, os estudos participantes dessa revisão apresentaram grandes diferenças metodológicas de acordo com o material biológico analisado, regiões espectrais analisadas, métodos de pré-processamentos e modelos de análise multivariada aplicados aos dados espectrais. Essa variabilidade contribuiu para as diferenças no desempenho dos modelos construídos para triagem de EM que pode ser observada pelas métricas apresentadas pelos artigos e descritas nesta revisão na Tabela 3. Além disso, a variação na métrica escolhida para avaliar a capacidade preditiva dos modelos dificulta a comparação entre os modelos, uma vez que alguns artigos utilizaram métricas como: sensibilidade e especificidade<sup>23,24</sup>, outros utilizaram a variância explicada entre os dados<sup>21,22</sup>, alguns também utilizaram a AUC<sup>22,23</sup> e outros realizaram apenas análises estatísticas pela comparação visual de bandas espectrais.<sup>25,26</sup>

No entanto, foi possível inferir que a escolha do biofluido e do *pipeline* analítico adotados exercem influência direta sobre o desempenho preditivo dos modelos de aprendizado de máquina. Estudos com plasma e soro apresentaram métricas de desempenho mais robustas, possivelmente devido à maior estabilidade do sangue e pela sua composição que reflete biomarcadores sistêmicos.<sup>21</sup> Por outro lado, análises baseadas em líquido cefalorraquidiano obtiveram alta capacidade para discriminação espectral em espectros de controle vs. EM<sup>22</sup>, mas essa metodologia

possui limitações relacionadas à coleta invasiva e ao tamanho reduzido das amostras, aumentando o risco de *overfitting*.

Além disso, há uma grande divergência entre as regiões espectrais utilizadas para construção dos modelos. No entanto, regiões espectrais associadas a lipídios (2950–2850  $\text{cm}^{-1}$ ) e proteínas (1650–1550  $\text{cm}^{-1}$ ) foram recorrentes, demonstrando que há uma diferença entre essas biomoléculas em pacientes com EM, o que pode ser explicado pela fisiopatologia da EM.<sup>10</sup> A ausência de consenso quanto à faixa espectral ideal para estudo da EM evidencia a necessidade de padronização de metodologias de aquisição espectral. Ademais, existe uma lacuna quanto aos tipos de pré-processamento espectrais necessários para uma melhor qualidade espectral. Nessa revisão, observou-se que estudos que aplicaram pré-processamentos como correção de linha de base, derivadas e normalização em conjunto com algoritmos de aprendizado de máquina supervisionado (RF, PLS-DA LDA) apresentaram melhor desempenho, reforçando a importância de *pipelines* analíticos robustos para uma melhor qualidade espectral e, consequentemente, melhor capacidade preditiva dos modelos.

Em comparação aos estudos incluídos nesta revisão, que estudaram a potencialidade da MIR para triagem de EM, estudos presentes na literatura que utilizaram a mesma técnica para outras doenças neurodegenerativas, como Alzheimer e Parkinson<sup>27,28,29</sup>, mostram padrões semelhantes de alteração em bandas lipídicas e proteicas, indicando que a MIR é capaz de refletir processos inflamatórios e neurodegenerativos. No entanto, ambos estudos apresentam pequeno grupo amostral, evidenciando a

necessidade de validações multicêntricas e maior número amostral para tornar os modelos confiáveis e interpretáveis.

Em termos de qualidade metodológica, a avaliação pelo QUADAS-2 indicou risco de viés moderado, principalmente na seleção de participantes e clareza na apresentação de resultados. Além disso, a falta de padronização na aquisição espectral e a ausência de validação externa foram limitações recorrentes.

Assim, recomenda-se que estudos futuros sobre a aplicabilidade da MIR para triagem de EM utilizem metodologias mínimas de padronização de acordo com a literatura. Por isso, preferencialmente deve-se adotar como biofluido o sangue ou suas frações (soro ou plasma) e regiões espectrais que compreendam as regiões de  $1700\text{--}900\text{cm}^{-1}$  e  $3000\text{--}2800\text{ cm}^{-1}$ . Ademais, é importante que tenha um detalhamento da seleção dos participantes, juntamente aos critérios de inclusão e exclusão adotados, descrição da aquisição espectral e dos pré-processamentos aplicados e *pipeline* da construção dos modelos de aprendizado de máquina com o uso de validação cruzada e validação externa. A implementação desses critérios permitirá maior comparabilidade entre estudos e contribuirá para consolidar a espectroscopia MIR como ferramenta de triagem clínica para a esclerose múltipla.

## 5. Conclusão

A presente revisão bibliométrica permitiu evidenciar o potencial da espectroscopia na região do MIR, associada a métodos de análise multivariada e aprendizado de máquina, como uma ferramenta promissora para triagem clínica da esclerose múltipla. Apesar do número

ainda reduzido de publicações que utilizam essa técnica para triagem da EM, os artigos analisados demonstram que diferentes biofluidos, técnicas espectrais e algoritmos de análise multivariada e aprendizado de máquina têm sido utilizados para discriminar amostras de indivíduos com EM de amostras de indivíduos saudáveis, sem comprometimento neurológico.

No entanto, a ausência de padronização entre os estudos, tanto em relação às metodologias de aquisição espectral quanto as análises estatísticas, ainda limita a reprodutibilidade e a comparabilidade entre os trabalhos. Nesse contexto, destaca-se a importância de investigações futuras adotarem diretrizes mínimas de padronização, incluindo: utilização do sangue ou suas frações (soro ou plasma) como biofluido; regiões espectrais que compreendam as regiões de  $1700\text{--}900\text{cm}^{-1}$  e  $3000\text{--}2800\text{ cm}^{-1}$ ; *pipeline* da construção dos modelos de aprendizado de máquina com o uso de validação cruzada e validação externa. Além de aumentar o número amostral, para que seja possível validar e consolidar a espectroscopia MIR como uma alternativa não invasiva, acessível e eficaz no apoio ao diagnóstico precoce da EM. Essa padronização será essencial para que a técnica avance da pesquisa para a prática clínica, contribuindo para diagnósticos mais rápidos e precisos.

## 6. Declaração de conflito de interesses

Os autores declaram que **não possuem conflitos de interesse** relacionados a este estudo.

## 7. Referências

1. CALLERY, E. L.; ROWBOTTOM, A. W. **Vibrational spectroscopy and multivariate analysis techniques in the clinical immunology laboratory: a review of current applications and requirements for diagnostic use.** *Applied Spectroscopy Reviews*, 1–30, 4 ago. 2021.  
<https://doi.org/10.1080/05704928.2021.1958337>
2. NASCIMENTO, M. H. C. et al. **Noninvasive diagnostic for COVID-19 from saliva biofluid via FTIR spectroscopy and multivariate analysis.** *Analytical Chemistry*, 94(5), 2425–2433, 25 jan. 2022.  
<https://doi.org/10.1021/acs.analchem.1c04162>
3. RAPOSO-NETO, J. J. et al. **Near-infrared spectroscopy with supervised machine learning as a screening tool for neutropenia.** *Journal of Personalized Medicine*, 14(1), 9, 21 dez. 2023.  
<https://doi.org/10.3390/jpm14010009>
4. KRAFT, C. et al. **Identification of B and T cells in human spleen sections by infrared microspectroscopic imaging.** *Cytometry Part A*, 64(2), 53–61, abr. 2005.  
<https://doi.org/10.1002/cyto.a.20117>
5. PALUSZKIEWICZ, C. et al. **Saliva as first-line diagnostic tool: a spectral challenge for identification of cancer biomarkers.** *Journal of Molecular Liquids*, 307, 2020.  
<https://doi.org/10.1016/j.molliq.2020.112961>
6. GULEKEN, Z. et al. **Characterization of COVID-19 infected pregnant women sera using laboratory indexes, vibrational spectroscopy and machine learning classifications.** *Talanta*, 237, 122916, jan. 2022.  
<https://doi.org/10.1016/j.talanta.2021.122916>
7. GIAMOUGIANNIS, P. et al. **Detection of ovarian cancer ( $\pm$  neo-adjuvant chemotherapy effects) via ATR-FTIR spectroscopy: comparative analysis of blood and urine biofluids in a large patient cohort.** *Analytical and Bioanalytical Chemistry*, 413(20), 5095–5107, 1 jul. 2021.  
<https://doi.org/10.1007/s00216-021-03472-8>
8. MARTINS, M. S. et al. **A machine learning approach towards endometriosis screening using infrared spectra of urine.** *Clinics*, 80, 100760, 6 set. 2025.  
<https://doi.org/10.1016/j.clinsp.2025.100760>
9. MACHADO, S. et al. **Recomendações esclerose múltipla.** São Paulo: Editora Omnifarma, 2012. 13 p. Disponível em: <https://pt-static.z-dn.net/files/dfa/306fd2fc1ff22a1408590c6e6e711b4c.pdf>. Acesso em: 10 ago. 2025
10. THOMPSON, A. et al. **Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria.** *Lancet Neurology*, 17, 162–173, 2018.  
[https://doi.org/10.1016/S1474-4422\(17\)30470-2](https://doi.org/10.1016/S1474-4422(17)30470-2)
11. MCDONALD, W.; COMPSTON, U. et al. **Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis.** *Annals of Neurology*, 50(1), 121–127, 2001.  
<https://doi.org/10.1002/ana.1032>
12. BRINER, R. B.; DENYER, D. **Systematic review and evidence synthesis as a practice and scholarship tool.** In: *The Oxford Handbook of Evidence-Based Management*. Oxford: Oxford University Press, 2012.  
<https://doi.org/10.1093/oxfordhb/9780199763986.013.0007>

13. BROADUS, R. N. **Toward a definition of “bibliometrics”**. *Scientometrics*, 12, 373–379, 1987.  
<https://doi.org/10.1007/BF02016680>
14. DIODATO, V. P.; GELLATLY, P. **Dictionary of Bibliometrics**. 11 jan. 2013.  
<https://doi.org/10.4324/9780203714133>
15. PRITCHARD, A. **Statistical bibliography or bibliometrics**. *Journal of Documentation*, 25, 348–349, 1969.
16. ARIA, M.; CUCCURULLO, C. **bibliometrix: an R-tool for comprehensive science mapping analysis**. *Journal of Informetrics*, 11(4), 959–975, nov. 2017.  
<https://doi.org/10.1016/j.joi.2017.08.007>
17. PAGANI, R. N.; KOVALESKI, J. L.; RESENDE, L. M. **Methodi ordinatio: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication**. *Scientometrics*, 105(3), 2109–2135, 12 set. 2015.  
<https://doi.org/10.1007/s11192-015-1744-x>
18. PAGE, M. J. et al. **The PRISMA 2020 statement: an updated guideline for reporting systematic reviews**. *British Medical Journal*, 372(71), 2021.  
<https://doi.org/10.1136/bmj.n71>
19. PAGE, M. J. et al. **PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews**. *British Medical Journal*, 372(160), 29 mar. 2021.  
<https://doi.org/10.1136/bmj.n160>
20. WHITING, P. F. **QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies**. *Annals of Internal Medicine*, 155(8), 529–536, 18 out. 2011. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>
21. KOŁODZIEJ, M. et al. **Spectral signature of multiple sclerosis. Preliminary studies of blood fraction by ATR FTIR technique**. *Biochemical and Biophysical Research Communications*, 593, 40–45, 1 jan. 2022.  
<https://doi.org/10.1016/j.bbrc.2022.01.046>
22. YONAR, D. et al. **Relapsing-remitting multiple sclerosis diagnosis from cerebrospinal fluids via Fourier transform infrared spectroscopy coupled with multivariate analysis**. *Scientific Reports*, 8(1), 18 jan. 2018.  
<https://doi.org/10.1038/s41598-018-19303-3>
23. EL KHOURY, Y. et al. **Rapid discrimination of neuromyelitis optica spectrum disorder and multiple sclerosis using machine learning on infrared spectra of sera**. *International Journal of Molecular Sciences*, 23(5), 2791, 3 mar. 2022.  
<https://doi.org/10.3390/ijms23052791>
24. CROCCO, M. C. et al. **ATR-FTIR spectroscopy of plasma supported by multivariate analysis discriminates multiple sclerosis disease**. *Scientific Reports*, 13(1), 13 fev. 2023.  
<https://doi.org/10.1038/s41598-023-29617-6>
25. TOMEČKOVÁ, V. et al. **Experimental analysis of tear fluid and its processing for the diagnosis of multiple sclerosis**. *Sensors*, 23(11), 5251, 1 jan. 2023.  
<https://doi.org/10.3390/s23115251>
26. CHOO, L. et al. **Infrared spectroscopic characterisation of multiple sclerosis plaques in the human central nervous system**. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1182(3), 333–337, 20 out. 1993
27. DEPCIUCH, J. et al. **FTIR spectroscopy of cerebrospinal fluid reveals variations in the lipid: protein ratio at different stages of Alzheimer’s disease**.

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 68(1), 281–293, 1 jan. 2019.  
<https://doi.org/10.3233/JAD-181008>

28. OFFNER, G.; ANDRE, W.; SANDT, C.; et al. **Synchrotron-based infrared spectroscopy brings to light the structure of protein aggregates in neurodegenerative diseases.** Reviews in Analytical Chemistry, 33, 231–243, 2014.  
<https://doi.org/10.1515/revac-2014-0016>

29. WANG, X. et al. **A study of Parkinson's disease patients' serum using FTIR spectroscopy.** Infrared Physics & Technology, 106, 103279–103279, 16 mar. 2020.  
<https://doi.org/10.1016/j.infrared.2020.103279>