

UNIVERSIDADE FEDERAL DE ALFENAS

DANIEL DE OLIVEIRA MIRANDA

**DESENVOLVIMENTO DE MÉTODOS COMPUTACIONAIS PARA CONSTRUÇÃO
DE MODELOS DE PREDIÇÃO DA ATIVIDADE ANTIBACTERIANA DE ÓLEOS
ESSENCIAIS**

ALFENAS/MG

2025

DANIEL DE OLIVEIRA MIRANDA

**DESENVOLVIMENTO DE MÉTODOS COMPUTACIONAIS PARA CONSTRUÇÃO
DE MODELOS DE PREDIÇÃO DA ATIVIDADE ANTIBACTERIANA DE ÓLEOS
ESSENCIAIS**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Química pela Universidade Federal de Alfenas. Área de concentração: Química Orgânica.

Orientadora: Prof^a. Dr^a. Marisi Gomes Soares
Coorientador: Prof. Dr. João Luiz Baldim Zanin

ALFENAS/MG

2025

Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Central

Miranda, Daniel de Oliveira.

Desenvolvimento de métodos computacionais para construção de modelos de predição da atividade antibacteriana de óleos essenciais / Daniel de Oliveira Miranda. - Alfenas, MG, 2025.

54 f. : il. -

Orientador(a): Marisi Gomes Soares.

Dissertação (Mestrado em Química) - Universidade Federal de Alfenas, Alfenas, MG, 2025.

Bibliografia.

1. Lamiaceae. 2. Staphylococcus aureus. 3. Atividade biológica. 4. Aprendizado de máquina. 5. J48. I. Soares, Marisi Gomes, orient. II. Título.

DANIEL DE OLIVEIRA MIRANDA

**DESENVOLVIMENTO DE MÉTODOS COMPUTACIONAIS PARA
CONSTRUÇÃO DE MODELOS DE PREDIÇÃO DA ATIVIDADE
ANTIBACTERIANA DE ÓLEOS ESSENCIAIS**

A Presidente da banca examinadora abaixo assina a aprovação da Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Química pela Universidade Federal de Alfenas. Área de concentração: Química Orgânica

Aprovado em: 24 de fevereiro de 2024.

Profa. Dra. Marisi Gomes Soares
Presidente da Banca Examinadora
Instituição: Universidade de Alfenas

Prof. Dr. Nerilson Marques Lima
Instituição: Universidade de Alfenas

Profa. Dra. Andreia Pereira Matos
Instituição: Universidade Federal de São Carlos



Documento assinado eletronicamente por **Marisi Gomes Soares, Professor do Magistério Superior**, em 07/04/2025, às 13:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1457603** e o código CRC **1347122D**.

Dedico este trabalho à Química, que desvela os segredos ocultos nas moléculas e nos ensina a linguagem da natureza, e ao essencial, que é invisível aos “óleos”, mas perceptível ao coração.

AGRADECIMENTOS

A conclusão desta dissertação representa não apenas o encerramento de uma etapa acadêmica, mas a realização de um sonho construído com apoio, incentivo e parceria de muitas pessoas, às quais sou profundamente grato. Sendo assim, agradeço, primeiramente, a Deus, pela força nos momentos de incerteza.

À minha orientadora, professora doutora Maris, minha sincera gratidão pela orientação dedicada, diálogo aberto e compromisso com minha formação. Aos membros da banca, professor doutor Nerilson e professora doutora Andreia, agradeço pelas valiosas contribuições e pelo tempo dedicado à leitura desta dissertação. Aos colegas e professores do Programa de Pós-Graduação em Química da Universidade Federal de Alfenas (UNIFAL-MG), agradeço pela convivência enriquecedora e pelas trocas de conhecimento – em especial às professoras Daniele D. e Daniella C., pela constante atuação ao lado da minha orientadora no Laboratório de Fitoquímica, Química Medicinal e Metabolômica da UNIFAL.

Sou grato aos amigos que estiveram ao meu lado nos momentos mais decisivos, oferecendo apoio, conselhos e presença: Roger, Davi M., João S., Sara L., Jean, Igor C., Karen, Cassiano, Tiago, Davi A., Gabriela M., Ana Caroline, Stephany, Jorge, Diogo, Jennifer, Lara, Paulo e Leandra.

À Caio César (em memória), cuja amizade e incentivo permanecem vivos em meu coração e lembranças.

À minha família, meu amor eterno e maior base. Em especial, à minha mãe, por seu apoio incondicional. Aos meus irmãos Daiane, Danilo e Daniella, por sempre estarem ao meu lado e aos meus primos Alan, Diego e Diene. Aos meus sobrinhos, Marcelo, Rodolfo, Maria Eduarda, Rhainna, Rhian, Jhonata e Luiz.

À Eliana, minha maior inspiração profissional e em quem me espelho em grande altruísmo.

Aos amigos Miller e Albert, agradeço a colaboração essencial em diversas etapas deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Agradeço à Pró-Reitoria de Pesquisa e Pós-Graduação da Universidade Federal de Alfenas, à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG – APQ-05218-23, APQ-02882-24, APQ-05607-24 e BPD-00760-22), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – 408115/2023-8, 316204/2021-8 e

406837/2021-0) e à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – 24/04606-5), pelo apoio financeiro que viabilizou a realização desta pesquisa.

A todos que, direta ou indiretamente, contribuíram para esta jornada e que sempre acreditaram em mim, especialmente nos momentos em que duvidei da minha própria capacidade, o meu mais sincero agradecimento.

RESUMO

A crescente resistência bacteriana aos antibióticos tem impulsionado a busca por alternativas terapêuticas eficazes e sustentáveis. Os óleos essenciais (OEs), metabólitos secundários de plantas, apresentam potencial antimicrobiano e vêm sendo estudados como agentes naturais contra patógenos resistentes. Este estudo teve como objetivo desenvolver métodos computacionais para a construção de modelos de predição da atividade antibacteriana de OEs da família Lamiaceae, utilizando abordagens de quimioinformática, aprendizado de máquina e estatística multivariada. A partir de uma revisão sistemática da literatura, foram coletados e organizados dados de cerca de 300 artigos científicos, estruturados em dois bancos de dados: (i) Banco de Composição Química, contendo informações sobre os constituintes dos OEs analisados, e (ii) Banco de Atividade Biológica, registrando ensaios antimicrobianos contra *Staphylococcus aureus*, um dos principais patógenos resistentes. Técnicas estatísticas foram aplicadas para a seleção de variáveis e identificação de padrões entre composição química e atividade antibacteriana. Na modelagem preditiva, foram testados algoritmos de aprendizado de máquina, incluindo Random Forest (RF), Redes Neurais Artificiais (RNA), Random Tree (RT), Naïve Bayes Tree (NBT) e J48 Decision Tree, sendo este último o mais eficaz, com 80,6% de acurácia na validação externa do modelo. Os resultados confirmam a aplicabilidade de modelos computacionais na predição da bioatividade de OEs, facilitando a seleção de compostos com potencial antimicrobiano. A integração entre quimioinformática e inteligência artificial otimiza a triagem de novos agentes terapêuticos, reduzindo tempo e custos experimentais, além de oferecer novas perspectivas para o combate à resistência antimicrobiana.

Palavras-chave: Lamiaceae; *Staphylococcus aureus*; atividade biológica; aprendizado de máquina; J48.

ABSTRACT

The increasing bacterial resistance to conventional antibiotics has driven the search for effective and sustainable therapeutic alternatives. Essential oils (EOs), secondary metabolites derived from plants, exhibit significant antimicrobial potential and have been extensively studied as natural agents against resistant pathogens. This study aimed to develop computational methods for constructing predictive models of the antibacterial activity of EOs from the Lamiaceae family, employing cheminformatics, machine learning, and multivariate statistical approaches. A systematic literature review was conducted, collecting and organizing data from approximately 300 scientific articles, which were structured into two main databases: (i) Chemical Composition Database, containing detailed information on the constituents of the analyzed EOs, and (ii) Biological Activity Database, documenting antimicrobial assays against *Staphylococcus aureus*, one of the main antibiotic-resistant pathogens. Statistical techniques were applied for variable selection and pattern identification between chemical composition and antibacterial activity. For predictive modeling, several machine learning algorithms were tested, including Random Forest (RF), Artificial Neural Networks (ANN), Random Tree (RT), Naïve Bayes Tree (NBT), and J48 Decision Tree, with the latter demonstrating the best performance, achieving 80.6% accuracy in external validation. The results confirm the applicability of computational models in predicting EOs bioactivity, facilitating the selection of compounds with antimicrobial potential. The integration of cheminformatics and artificial intelligence enhances the screening process for new therapeutic agents, reducing experimental time and costs while offering new perspectives for combating antimicrobial resistance.

Keywords: Lamiaceae; *Staphylococcus aureus*; biological activity; machine learning; J48.

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1 - Curadoria dos Dados e Geração dos Modelos de Predição | 28 |
| Figura 2 - Fluxo de trabalho utilizado na elaboração dos modelos de predição | 30 |
| Figura 3 - Árvore de Decisão J48 - Modelo de Predição | 42 |

LISTA DE QUADROS

| | |
|--|----|
| Quadro 1 – Estrutura e descrição do banco de dados de composição | 33 |
| Quadro 2 – Estrutura e descrição do banco de dados de atividade..... | 34 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 - Dados de teste curados – utilizados para elaboração dos modelos de predição | 36 |
| Tabela 2 - Dados de teste curados – utilizados para elaboração dos modelos de predição..... | 39 |
| Tabela 3 - Métricas de desempenho para modelos..... | 40 |
| Tabela 4 - Ajuste de hiperparâmetros ótimos do modelo de predição J48..... | 43 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|----------------------|---|
| ANN | Artificial Neural Networks (Redes Neurais Artificiais) |
| ANOVA | Analysis of Variance (Análise de Variância) |
| <i>B. cereus</i> | <i>Bacillus cereus</i> |
| BDA | Banco de Dados de Atividade |
| BDC | Banco de Dados de Composição |
| CAPES | Coordenação de Aperfeiçoamento de Pessoal de Nível Superior |
| CG-EM | Cromatografia Gasosa acoplada à Espectrometria de Massas |
| CLSI | Clinical and Laboratory Standards Institute (Instituto de Padrões Clínicos e Laboratoriais) |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| EUCAST | European Committee on Antimicrobial Susceptibility Testing (Comitê Europeu de Testes de Suscetibilidade Antimicrobiana) |
| IUPAC | União Internacional de Química Pura e Aplicada |
| J48 | Algoritmo de Árvore de Decisão baseado no C4.5 |
| KNIME | Konstanz Information Miner |
| MIC | Minimum Inhibitory Concentration (Concentração Inibitória Mínima) |
| NBT | Naïve Bayes Tree |
| OE | Óleo Essencial |
| PCA | Principal Component Analysis (Análise de Componentes Principais) |
| PLS | Partial Least Squares (Mínimos Quadrados Parciais) |
| QSAR | Quantitative Structure–Activity Relationship (Relação Estrutura-Atividade Quantitativa) |
| QSTR | Quantitative Structure-Toxicity Relationship (Relação Quantitativa Estrutura-Toxicidade) |
| RF | Random Forest (Floresta Randômica) |
| RT | Random Tree (Árvore Randômica) |
| SMILES | Simplified Molecular Input Line Entry System (Sistema Simplificado de Entrada Linear para Representação Molecular) |
| <i>P. aeruginosa</i> | <i>Pseudomonas aeruginosa</i> |
| <i>S. aureus</i> | <i>Staphylococcus aureus</i> |

SUMÁRIO

| | | |
|----------|--|-----------|
| 1 | INTRODUÇÃO | 15 |
| 1.1 | OBJETIVOS..... | 17 |
| 1.1.1 | Objetivo geral..... | 17 |
| 1.1.2 | Objetivos específicos..... | 17 |
| 2 | REVISÃO BIBLIOGRÁFICA | 18 |
| 2.1 | POTENCIAL ANTIBACTERIANO DE ÓLEOS ESSENCIAIS: APLICAÇÕES METABOLÔMICAS E MODELAGEM PREDITIVA | 18 |
| 2.2 | LAMIACEAE..... | 19 |
| 2.3 | <i>STAPHYLOCOCCUS AUREUS</i> | 21 |
| 2.4 | QUIMIOINFORMÁTICA NO DESENVOLVIMENTO DE MODELOS DE PREDIÇÃO DE ATIVIDADE BIOLÓGICA | 22 |
| 3 | METODOLOGIA..... | 25 |
| 3.1 | CRIAÇÃO DOS BANCOS DE DADOS..... | 25 |
| 3.1.1 | Construção e Organização dos Bancos de Dados | 25 |
| 3.1.2 | Coleta e Organização dos Dados | 25 |
| 3.1.3 | Estruturação, Filtros e Conversão de Unidades | 26 |
| 3.1.4 | Integração e Refinamento dos Dados..... | 27 |
| 3.1.5 | Curadoria e Pré-Processamento de Dados Químicos..... | 27 |
| 3.2 | CONSTRUÇÃO DOS MODELOS DE PREDIÇÃO | 28 |
| 3.2.1 | Abordagem de Classificação | 28 |
| 3.2.2 | Modelos de Predição..... | 29 |
| 3.2.3 | Validação do Modelo | 30 |
| 3.2.4 | Ajuste de Hiperparâmetros..... | 30 |
| 4 | RESULTADOS E DISCUSSÃO | 32 |
| 4.1 | COLETA DE DADOS E CONSTRUÇÃO DOS BANCOS DE DADOS | 32 |
| 4.2 | BANCO DE DADOS PRINCIPAL PARA MODELO DE PREDIÇÃO DE ATIVIDADE DE ÓLEOS ESSENCIAIS DE ESPÉCIES DA FAMÍLIA LAMIACEAE CONTRA <i>STAPHYLOCOCCUS AUREUS</i> | 35 |
| 4.2.1 | Seleção de atributos e critérios de desempenho para elaboração dos | |

| | |
|---|-----------|
| modelos a partir do Banco de Dados Principal..... | 35 |
| 4.2.2 Construção dos Modelos de Predição de Atividade Antibacteriana contra <i>Staphylococcus aureus</i> | 40 |
| 5 CONCLUSÃO..... | 45 |
| REFERÊNCIAS..... | 47 |
| APÊNDICE | 52 |

1 INTRODUÇÃO

Os óleos essenciais (OEs) são complexas combinações de substâncias voláteis derivadas de plantas, armazenadas de forma específica em diversas partes, como folhas, flores, ramos, frutos e raízes. A literatura descreve inúmeras aplicações para os OEs, sendo amplamente utilizados na conservação de alimentos, nas indústrias farmacêutica, cosmética e de perfumes, além de terem relevância na medicina, aromaterapia, microbiologia, agricultura, pecuária, entre outras áreas. Dentre suas diversas propriedades, destaca-se o significativo potencial antimicrobiano (Pezantes-Orellana *et al.*, 2024). Essa atividade contra diversos patógenos é atribuída à natureza, composição e concentração dos seus componentes, estando diretamente relacionada à sinergia e proporção entre eles (Balim *et al.*, 2018).

A busca por substâncias terapêuticas mais eficazes, com menor toxicidade e maior especificidade, tem impulsionado intensamente os estudos voltados à descoberta de novos fármacos. Estima-se que aproximadamente 25% dos medicamentos utilizados em países industrializados sejam derivados, direta ou indiretamente, de produtos naturais (Newman; Cragg, 2012). Um dos principais fatores que reforçam essa busca é o avanço da resistência bacteriana aos antibióticos atualmente disponíveis, o que tem comprometido a eficácia dos tratamentos convencionais (Brandt *et al.*, 2014; Mitra *et al.*, 2023; Gelatti *et al.*, 2009).

De acordo com a Organização Mundial da Saúde (OMS), essa resistência representa uma ameaça crescente. Dados de 76 países revelam que 42% das cepas de *Escherichia coli* são resistentes a cefalosporinas de terceira geração, e 35% das cepas de *Staphylococcus aureus* apresentam resistência à metilina. Em 2020, cerca de 20% das infecções do trato urinário causadas por *E. coli* mostraram sensibilidade reduzida a antibióticos comuns, como ampicilina, co-trimoxazol e fluoroquinolonas (World Health Organization, 2023). Esse cenário, que dificulta o tratamento eficaz de infecções rotineiras, evidencia a urgência na busca por novas alternativas terapêuticas.

Nesse contexto, os OEs emergem como agentes antibacterianos promissores. Sua complexa composição química confere múltiplos mecanismos de ação, dificultando o surgimento de resistência bacteriana, já que os compostos atuam por vias distintas (Carvalho, 2023). Assim, a identificação e caracterização de seus constituintes são etapas essenciais para a exploração do potencial terapêutico desses metabólitos naturais.

Espécies vegetais pertencentes a famílias conhecidas pela produção de OEs, como Lamiaceae, Myrtaceae, Asteraceae, Rutaceae e Apiaceae, têm demonstrado eficácia contra uma variedade de microrganismos patogênicos, incluindo fungos, vírus e bactérias (Swamy; Akhtar;

Sinniah, 2016). Dentre essas, a família Lamiaceae destaca-se por reunir diversas espécies aromáticas extensivamente estudadas por suas propriedades antimicrobianas.

Para além da identificação química, a utilização de métodos computacionais tem se tornado indispensável na análise de produtos naturais, especialmente frente à imensa quantidade de dados gerados por técnicas como Cromatografia Gasosa acoplada à Espectrometria de Massas. As análises metabolômicas permitem uma abordagem holística dos metabólitos presentes nas amostras vegetais, favorecendo a compreensão de suas atividades biológicas. A construção de modelos preditivos e classificatórios com base nesses dados representa, portanto, uma estratégia valiosa para guiar a descoberta de novos compostos bioativos com potencial terapêutico.

Além de facilitar a triagem de substâncias promissoras, esses modelos computacionais viabilizam a quantificação e compreensão das combinações de compostos essenciais à eficácia antibacteriana, por meio de abordagens multivariadas. Considerando a complexidade das matrizes orgânicas, é essencial o uso de ferramentas avançadas, como Aprendizado de Máquina (*Machine Learning*) e Estatística Multivariada, que permitem a interpretação robusta e precisa dos dados (Chagas-Paula *et al.*, 2015).

Diante da ampla aplicação dos OEs como agentes antimicrobianos e da disponibilidade crescente de dados na literatura, este trabalho tem como objetivo principal desenvolver modelos computacionais capazes de prever a atividade antibacteriana de OEs provenientes de espécies da família Lamiaceae, uma das mais relevantes entre as produtoras de compostos voláteis bioativos. A atividade antibacteriana foi escolhida como foco de análise, tendo em vista que microrganismos como *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *E. coli*, *Bacillus cereus* e *Salmonella* spp. estão fortemente associados à resistência antimicrobiana e são responsáveis por infecções e intoxicações alimentares em humanos (Mitra *et al.*, 2023; Fischbach; Walsh, 2009).

Os modelos de predição desenvolvidos são de caráter classificatório, utilizando algoritmos de Aprendizado de Máquina e Métodos Estatísticos Multivariados como PCA (Análise de Componentes Principais), PLS (Mínimos Quadrados Parciais), árvores de decisão e clustering.

1.1 OBJETIVOS

1.1.1 Objetivo geral

Desenvolver métodos computacionais para elaboração de modelos de predição de atividade antibacteriana de OEs de espécies da família botânica Lamiaceae frente ao microrganismo patogênico *S. aureus*.

1.1.2 Objetivos específicos

- a) realizar pesquisa bibliográfica abrangente baseada na composição de OEs obtidos a partir de plantas que tenham seus compostos identificados por índice de Kovats ou Índice de Retenção e que tenham sido testados contra bactérias patogênicas gerando unidades de Concentração Inibitória Mínima (MIC) ou Zona de Inibição;
- b) organizar os dados com os resultados da busca bibliográfica de acordo com a composição e a atividade antibacteriana testada;
- c) criar bancos de dados contendo informações dos componentes identificados e atividades biológicas respectivas a cada óleo essencial;
- d) avaliar a possibilidade de ajuste dos bancos de dados obtidos removendo outliers, a fim de melhorar a capacidade classificatória ou preditiva dos modelos criados;
- e) identificar os compostos mais relevantes que compõem os OEs através de modelos de árvore de decisão (J48);
- f) testar, quando disponíveis, os compostos apontados como biomarcadores pelos modelos a fim de averiguar suas atividades como agentes antibacterianos.

2 REVISÃO BIBLIOGRÁFICA

2.1 POTENCIAL ANTIBACTERIANO DE ÓLEOS ESSENCIAIS: APLICAÇÕES METABOLÔMICAS E MODELAGEM PREDITIVA

A descoberta de novos agentes antibacterianos tem sido inversamente proporcional ao surgimento de mecanismos de resistência aos antibióticos. Desde 1960 até a década de 2000, a ausência de novos antibióticos com mecanismos de ação ou classes estruturais diferenciadas intensificou o estudo dos genes de resistência. Esse cenário levou Fischbach e Walsh (2009), a cunhar o termo "lacuna de inovação", indicando a falta de novas classes de moléculas, limitando-se a modificações em compostos existentes.

Diante dessa limitação, a busca por alternativas eficazes impulsionou o interesse por produtos naturais, especialmente os derivados de plantas. Nesse contexto, os OEs surgem como uma fonte promissora de compostos antibacterianos (Swamy; Akhtar; Sinniah, 2016). Sua complexidade e diversidade química favorecem múltiplos mecanismos de ação, dificultando o desenvolvimento de resistência bacteriana.

Com o aumento do interesse no uso terapêutico de fitoterápicos, tanto extratos fixos quanto voláteis, métodos computacionais vêm sendo empregados para identificar substâncias promissoras. Entre essas abordagens, a Análise Estatística Multivariada e o Aprendizado de Máquina têm demonstrado potencial na anotação de compostos bioativos e predição de suas atividades biológicas (Baldim *et al.*, 2017; Zhang *et al.*, 2014). Essas ferramentas possibilitam uma compreensão mais aprofundada das interações entre os componentes dos OEs e suas propriedades antibacterianas ao permitir a identificação de padrões de correlação entre variáveis químicas e atividades biológicas. Por meio de técnicas como Análise de Componentes Principais (PCA), Mínimos Quadrados Parciais (PLS) e algoritmos supervisionados de classificação, como árvores de decisão e Random Forest, é possível reduzir a dimensionalidade dos dados, evidenciar agrupamentos, e apontar quais compostos ou combinações são mais relevantes na resposta antibacteriana observada.

Apesar dos avanços na caracterização química dos OEs, ainda há uma lacuna na literatura sobre análises metabolômicas abrangentes que considerem múltiplas espécies e suas composições químicas complexas. A metabolômica, uma abordagem detalhada para o estudo de metabólitos de diversos organismos, incluindo plantas, bactérias e fungos, tem se mostrado uma ferramenta valiosa para esse tipo de investigação (Canuto *et al.*, 2017). Assim, sua aplicação pode contribuir significativamente para a compreensão dos mecanismos de ação dos

óleos essenciais, uma vez que a metabolômica permite a detecção, quantificação e comparação de perfis metabólicos completos, mesmo em amostras complexas. Ao integrar dados químicos obtidos por técnicas como Cromatografia Gasosa acoplada à Espectrometria de Massas (CG-EM), a metabolômica facilita a identificação de compostos majoritários e minoritários, e possibilita a correlação entre a composição química e os efeitos biológicos observados. Dessa forma, torna-se viável reconhecer padrões moleculares específicos associados à atividade antibacteriana, bem como possíveis interações sinérgicas entre os constituintes voláteis.

Estudos anteriores já demonstraram a importância da variação química entre diferentes espécies e dentro da mesma espécie. Por exemplo, uma pesquisa com OEs de *Ocimum basilicum* identificou compostos com atividade antimicrobiana relevante, especialmente contra *B. cereus*, um microrganismo associado a intoxicações alimentares (Baldim *et al.*, 2018). Esses achados reforçam a necessidade de estudos mais amplos para correlacionar a composição química dos OEs com sua eficácia antibacteriana.

Diante do crescente uso dos OEs como agentes antibacterianos e da ampla disponibilidade de informações na literatura, torna-se essencial adotar métodos que permitam uma classificação e predição mais sistemática dessas substâncias. A identificação das principais características estruturais e físico-químicas de seus componentes pode facilitar a construção de modelos de predição capazes de indicar quais OEs apresentam maior potencial antibacteriano. Dessa forma, esses modelos não apenas otimizam a seleção de candidatos promissores para testes laboratoriais, como também direcionam novas estratégias para o desenvolvimento de agentes antimicrobianos eficazes.

2.2 LAMIACEAE

A família Lamiaceae é amplamente reconhecida por sua diversidade e importância ecológica, abrangendo aproximadamente 236 gêneros e cerca de 6.000 espécies distribuídas globalmente. Muitas dessas espécies são valorizadas devido à sua capacidade de produzir metabólitos secundários, especialmente OEs, os quais desempenham um papel fundamental na defesa da planta e apresentam diversas propriedades bioativas (Ebadollahi; Ziaee; Palla, 2020). Os OEs derivados das Lamiaceae são compostos predominantemente por monoterpenos e sesquiterpenos, o que lhes confere um amplo espectro de atividades biológicas. Essas substâncias possuem propriedades antibacterianas, antifúngicas, antioxidantes, antivirais e até mesmo anticancerígenas, o que tem levado a um crescente interesse por seu uso nas indústrias farmacêutica, cosmética e alimentícia (Ali *et al.*, 2024; Elbouzidi *et al.*, 2024).

A presença de compostos bioativos em OEs da família Lamiaceae tem sido amplamente estudada devido ao seu potencial terapêutico e farmacológico. Estudos recentes destacam que espécies como *Thymus vulgaris*, *Rosmarinus officinalis*, *Salvia officinalis* e *Mentha piperita* apresentam um conjunto variado de metabólitos com atividades antimicrobianas, anti-inflamatórias e neuroprotetoras (Skrovankova; Mlcek, 2025).

- Atividade Antibacteriana: Diversas espécies de Lamiaceae têm demonstrado ação contra bactérias patogênicas, incluindo cepas resistentes a antibióticos. O óleo essencial de *Origanum vulgare* e *Thymus vulgaris*, por exemplo, exibe forte atividade antibacteriana devido à presença de carvacrol e timol, que atuam na desestabilização da membrana celular dos microrganismos (Sar *et al.*, 2023).

- Atividade Antioxidante: Compostos voláteis como 1,8-cineol, carnosol, α -terpineol, borneol e ácido rosmarínico, comumente presentes nos óleos essenciais de *Salvia officinalis* e *Rosmarinus officinalis*, desempenham um papel relevante na neutralização de radicais livres. Esses compostos têm sido amplamente estudados por sua capacidade de atuar como agentes antioxidantes, contribuindo para a prevenção de danos oxidativos em células biológicas e mostrando potencial na prevenção de doenças neurodegenerativas e no combate ao estresse oxidativo (Ali *et al.*, 2024; Bejenaru *et al.*, 2024).

- Atividade Antifúngica: o óleo essencial de *Lavandula angustifolia* tem sido avaliado por sua eficácia contra fungos patogênicos, incluindo espécies do gênero *Candida*, tornando-se um potencial adjuvante para tratamentos antifúngicos convencionais (D; K; F, 2024; Viswanathan; Krishnan, 2024).

A utilização de OEs dessas espécies vai além da farmacologia. Seu emprego na indústria cosmética como antioxidantes naturais tem crescido, principalmente em formulações para cuidados com a pele. Além disso, a indústria alimentícia tem explorado essas substâncias como conservantes naturais, devido às suas propriedades antimicrobianas (Ali *et al.*, 2024).

Dentre os gêneros mais estudados dentro da família Lamiaceae, destacam-se *Lavandula*, *Mentha*, *Origanum*, *Thymus* e *Rosmarinus*, amplamente reconhecidos por suas propriedades terapêuticas e aplicações industriais. O gênero *Lavandula* é frequentemente utilizado na aromaterapia e na indústria cosmética devido à sua ação calmante e aos efeitos positivos na qualidade do sono e na redução da ansiedade, conforme demonstrado por ensaios clínicos com

aromaterapia em pacientes com doenças cardiovasculares (Chien *et al.*, 2023). Já as espécies do gênero *Mentha*, como a hortelã-pimenta (*Mentha piperita*), são exploradas pelas indústrias alimentícia e farmacêutica por suas propriedades digestivas e antissépticas, sendo eficazes no alívio de sintomas associados à síndrome do intestino irritável (Khosravi *et al.*, 2020). Os gêneros *Origanum* e *Thymus* se destacam pelo elevado potencial antimicrobiano e antioxidante de seus óleos essenciais, atribuídos à presença de compostos como carvacrol e timol, com aplicação como conservantes naturais (González *et al.*, 2024). Por fim, o gênero *Rosmarinus* tem despertado interesse em estudos relacionados à neuroproteção, especialmente pela atuação de compostos voláteis no combate ao estresse oxidativo e na modulação de neurotransmissores, contribuindo para a melhora da memória e da função cognitiva (Al-Kuraishi *et al.*, 2024).

A crescente busca por alternativas naturais para combater doenças infecciosas e o aumento da resistência bacteriana tem intensificado os estudos com a família Lamiaceae, tornando seus OEs alvos promissores para o desenvolvimento de novos fármacos e cosméticos de base natural.

2.3 *Staphylococcus aureus*

O uso de OEs como agentes antibacterianos representa uma alternativa promissora e sustentável para enfrentar a crescente resistência bacteriana aos antimicrobianos sintéticos. Entre os diversos patógenos, *S. aureus* destaca-se por sua relevância clínica, sendo responsável por infecções que variam de lesões cutâneas superficiais a infecções sistêmicas graves. Sendo assim, a escolha de *S. aureus* como objetivo deste projeto se deve à sua recorrência nos estudos coletados: entre 300 artigos analisados, cerca de 135 mencionam essa bactéria, tornando-a um alvo bem documentado. Além disso, *S. aureus* integra a microbiota da pele e pode se tornar patogênico em situações de comprometimento da barreira cutânea ou imunidade reduzida. Essa bactéria é uma das principais causas de infecções nosocomiais e comunitárias, associando-se a altos índices de morbidade e mortalidade, além de apresentar significativa resistência antimicrobiana (Gelatti *et al.*, 2009).

Os OEs extraídos de plantas medicinais e aromáticas têm demonstrado atividade antimicrobiana contra *S. aureus*, incluindo cepas resistentes à metilina (MRSA). Estudos *in vitro*, como os de Penteado *et al.* (2021), indicam que OEs de tomilho (*Thymus vulgaris*), orégano (*Origanum vulgare*) e alecrim-pimenta (*Lippia sidoides*) contêm compostos majoritários, como timol, carvacrol e para-cimeno, altamente eficazes na inibição do crescimento de *S. aureus* (Penteado *et al.*, 2021). Uma revisão de Swamy *et al.* (2016) analisou

a atividade antibacteriana de OEs com base em 72 referências, evidenciando sua eficácia contra diversas cepas bacterianas, incluindo *E. coli* e *P. aeruginosa*.

Os mecanismos de ação dos OEs incluem a alteração da permeabilidade da membrana celular dos patógenos, provocando vazamento de constituintes intracelulares e interrupção do metabolismo celular (Swamy; Akhtar; Sinniah, 2016).

A escolha por tal microrganismo se justifica, conforme mencionado anteriormente, por sua relevância clínica e pela disponibilidade de dados que possibilitam uma análise robusta da eficácia dos OEs como agentes antibacterianos. Considerando a crescente resistência dessa bactéria aos antibióticos tradicionais, a exploração de alternativas naturais torna-se urgente. Assim, esta pesquisa não apenas possibilitará contribuições para o combate às infecções causadas por *S. aureus*, mas também poderá abrir caminho para estudos da aplicação de OEs no controle de outros microrganismos, ampliando seu impacto na saúde pública e na segurança alimentar.

2.4 QUIMIOINFORMÁTICA NO DESENVOLVIMENTO DE MODELOS DE PREDIÇÃO DE ATIVIDADE BIOLÓGICA

A quimioinformática é uma ciência interdisciplinar que combina conhecimentos da química e da ciência da computação para analisar, modelar e interpretar grandes volumes de dados químicos e biológicos. Desde sua concepção na década de 1960, a área tem evoluído significativamente, passando da simples representação e manipulação de estruturas químicas para a incorporação de técnicas avançadas de aprendizado de máquina, inteligência artificial e análise estatística multivariada. Essa evolução tem permitido a descoberta de novos compostos bioativos, a modelagem de interações moleculares e a predição de propriedades químicas e biológicas, sendo uma ferramenta essencial na pesquisa farmacêutica, agroquímica e na química de produtos naturais (Alves *et al.*, 2017).

No contexto da predição de atividade biológica, os modelos QSAR (Quantitative Structure-Activity Relationship - Relação Estrutura-Atividade Quantitativa) são amplamente utilizados para correlacionar as estruturas químicas de compostos com suas propriedades biológicas por meio de descritores moleculares e algoritmos matemáticos. A construção de modelos QSAR envolve a coleta e o tratamento de dados químicos e biológicos, a seleção de descritores relevantes e a aplicação de métodos estatísticos ou de aprendizado de máquina para a criação de equações preditivas. Dependendo da abordagem utilizada, os modelos podem ser classificados em globais, que consideram todo o conjunto de dados para gerar as equações

preditivas, ou locais, que analisam subconjuntos específicos de moléculas com alta similaridade estrutural ou mecanismos de ação semelhantes (Dhane *et al.*, 2022; Diéguez-Santana; González-Díaz, 2023).

Embora os modelos QSAR sejam amplamente empregados na quimioinformática, novas estratégias vêm sendo exploradas para superar suas limitações e ampliar o escopo da predição de bioatividade. Uma dessas estratégias envolve o uso de dados metabolômicos, que permitem uma abordagem mais abrangente e holística na análise de sistemas químicos e biológicos. A metabolômica é um campo da química analítica que estuda o conjunto de metabólitos presentes em organismos vivos, fornecendo informações detalhadas sobre os processos metabólicos e suas interações com compostos exógenos, como fármacos e substâncias naturais. Esses dados são obtidos por meio de técnicas como espectrometria de massas e ressonância magnética nuclear, gerando perfis químicos complexos que podem ser analisados com ferramentas computacionais para identificar padrões bioativos e biomarcadores metabólicos relevantes (Dunn; Ellis, 2005; Pilon *et al.*, 2020).

A aplicação da quimioinformática na análise metabolômica permite a construção de modelos predição capazes de correlacionar assinaturas metabólicas com atividades biológicas, sem depender exclusivamente da estrutura molecular dos compostos individuais. Essa abordagem tem sido particularmente útil na análise de misturas complexas, como extratos vegetais e fluidos biológicos, onde a atividade biológica resulta da interação entre múltiplos metabólitos. Além disso, a metabolômica integrada à inteligência artificial tem possibilitado avanços significativos na predição de bioatividade, por meio da aplicação de técnicas como análise de componentes principais (PCA), máquinas de vetores de suporte (SVM), redes neurais artificiais (ANNs) e algoritmos de agrupamento, que permitem extrair padrões ocultos e melhorar a acurácia das previsões (Maraschin *et al.*, 2017; Wang *et al.*, 2016).

Além da modelagem QSAR e da análise metabolômica, a quimioinformática tem sido aplicada em diversas outras frentes na pesquisa química e biológica. Entre suas principais aplicações, destacam-se a predição de propriedades físico-químicas, como solubilidade, permeabilidade e estabilidade química, que são fundamentais para a triagem de novos fármacos; a modelagem de interações proteína-ligante, utilizando técnicas de docking molecular para prever como compostos químicos interagem com alvos biológicos; e a quimiogenômica, que busca relacionar características químicas dos compostos com seus efeitos sobre diferentes alvos genéticos. Além disso, a quimioinformática também tem sido utilizada na predição de toxicidade, por meio de abordagens QSTR (Quantitative Structure-Toxicity Relationship), que avaliam o risco toxicológico de substâncias químicas antes de sua aplicação industrial ou

farmacêutica, auxiliando na identificação de potenciais efeitos adversos e garantindo maior segurança na introdução de novos compostos químicos no mercado (Menezes; Scotti; Scotti, 2024).

A importância da quimioinformática no desenvolvimento de modelos de predição tem sido amplamente reconhecida na literatura. Estudos recentes, como o de Menezes *et al.* (2024), reforçam a necessidade de aprimorar a construção e a validação desses modelos, explorando novas abordagens de aprendizado de máquina para aumentar a precisão das previsões e reduzir a dependência de métodos tradicionais. Além disso, o estudo enfatiza a importância da explicabilidade e interpretabilidade dos modelos, garantindo que suas predições sejam cientificamente justificáveis e compreensíveis, e discute a necessidade de diretrizes éticas no uso da inteligência artificial em modelagens computacionais.

Dessa forma, a quimioinformática tem se consolidado como uma ferramenta indispensável para a descoberta e otimização de novos compostos bioativos. A evolução das técnicas computacionais e a integração de dados químicos e biológicos têm permitido avanços significativos na pesquisa farmacêutica, na química de produtos naturais e na biotecnologia. A incorporação de abordagens metabolômicas e o uso de inteligência artificial representam um novo paradigma na modelagem da bioatividade de compostos, oferecendo alternativas mais robustas e abrangentes para a predição de propriedades químicas e biológicas. Com a contínua evolução das metodologias computacionais e o aumento da disponibilidade de grandes volumes de dados químicos, espera-se que a quimioinformática desempenhe um papel ainda mais central na pesquisa e desenvolvimento de novos fármacos, agroquímicos e materiais bioativos, contribuindo para o avanço da ciência e da tecnologia de forma cada vez mais sustentável e com propostas inovadoras.

3 METODOLOGIA

3.1 CRIAÇÃO DOS BANCOS DE DADOS

3.1.1 Construção e Organização dos Bancos de Dados

Uma das premissas deste estudo foi a construção de Bancos de Dados estruturados que permitissem a análise da relação entre a composição química de OEs e sua atividade antibacteriana. Para isso, foi realizado um levantamento sistemático de informações disponíveis na literatura científica, com a organização de um banco de dados de composição química dos OEs e sua atividade antimicrobiana contra diversos microrganismos. A partir dessa base consolidada, foi possível extrair um conjunto de dados específico para o desenvolvimento de modelos computacionais aplicados à predição da atividade antibacteriana.

3.1.2 Coleta e Organização dos Dados

A obtenção dos dados seguiu uma abordagem sistemática baseada na revisão da literatura científica e extração padronizada de informações de artigos publicados. O processo iniciou-se com a definição de critérios para a seleção dos estudos, visando garantir a relevância e confiabilidade dos dados coletados. Foram consultadas bases científicas amplamente reconhecidas, como Science Direct, SciELO, Periódicos CAPES, Scopus, PubChem, SpectraBase e NIST. A pesquisa bibliográfica utilizou palavras-chave específicas relacionadas aos óleos essenciais (OEs), sua composição química e atividade antimicrobiana, tais como: "essential oils", "chemical composition", "volatile compounds", "antibacterial activity", "Lamiaceae", "GC-MS" (Gas Chromatography-Mass Spectrometry), "natural products", "bioactive compounds", entre outras. Para refinar as buscas e maximizar a recuperação de artigos relevantes, foram utilizados operadores booleanos como AND, OR e NOT — por exemplo: "essential oils" AND "antibacterial activity" AND "Lamiaceae", ou "volatile compounds" OR "terpenes" AND "bioactivity".

Após a busca inicial, os estudos selecionados passaram por uma triagem baseada na leitura dos títulos e resumos, sendo excluídos aqueles que não apresentavam dados quantitativos ou qualitativos sobre a composição química dos óleos essenciais (OEs) ou que não continham ensaios padronizados de atividade antimicrobiana. Foram considerados apenas os artigos que forneciam informações detalhadas sobre a metodologia experimental utilizada para a

identificação dos compostos químicos, bem como sobre os testes microbiológicos realizados. Entre os critérios de inclusão, destacaram-se a utilização de ensaios antimicrobianos padronizados, como determinação da Concentração Inibitória Mínima (MIC), Concentração Bactericida Mínima (MBC), ensaio de difusão em disco (disk diffusion) e técnicas de microdiluição em caldo (broth microdilution), conforme diretrizes internacionalmente reconhecidas, como as do Clinical and Laboratory Standards Institute (CLSI) ou da European Committee on Antimicrobial Susceptibility Testing (EUCAST).

Os dados foram extraídos manualmente de cada artigo selecionado, utilizando uma abordagem estruturada para evitar erros de transcrição. Cada artigo foi analisado individualmente, e os dados foram registrados de forma tabular, garantindo a rastreabilidade das informações. Os OEs foram organizados conforme suas respectivas famílias botânicas, incluindo Lamiaceae, Myrtaceae, Rutaceae, Asteraceae, Fabaceae, Apiaceae, Zingiberaceae, entre outras. A atividade antimicrobiana foi registrada para diferentes microrganismos, como *S. aureus*, *E. coli*, *P. aeruginosa*, *Klebsiella pneumoniae*, *Enterococcus faecalis*, entre outros.

Para minimizar o viés e assegurar a precisão dos dados, foi realizada uma verificação cruzada das informações extraídas, comparando-as com os valores originais apresentados nos artigos. Além disso, foram consultados bancos de dados químicos para validar a nomenclatura dos compostos e seus identificadores estruturais, como SMILES (Sistema Simplificado de Entrada Linear para Representação Molecular) - Canônico e Isomérico.

3.1.3 Estruturação, Filtros e Conversão de Unidades

Os dados coletados foram organizados em dois bancos de dados principais: um contendo informações sobre a composição química dos OEs (Banco de Dados de Composição - BDC) e outro reunindo os dados de atividade antimicrobiana (Banco de Dados de Atividade - BDA). No banco de composição química, foram registradas as informações sobre os compostos presentes em cada óleo essencial, incluindo nomes IUPAC (União Internacional de Química Pura e Aplicada), SMILES canônico e isomérico, além da porcentagem relativa de cada composto. No banco de atividade antimicrobiana, foram armazenados dados referentes aos microrganismos testados, métodos utilizados (Zona de Inibição e Concentração Inibitória Mínima - MIC) e seus respectivos valores experimentais obtidos nos ensaios microbiológicos.

Para garantir a qualidade e padronização das informações, foram aplicados filtros rigorosos em ambos os bancos. No banco de composição química, foram removidos registros com dados incompletos ou inconsistentes, garantindo que ao menos 70% da composição

química total dos OEs estivesse identificada. No banco de atividade antimicrobiana, foram excluídos estudos que não apresentavam métodos padronizados ou que continham resultados incompatíveis com a análise quantitativa - essas exclusões foram unificadas para ambos os bancos de dados, ou seja, se uma exclusão era feita em um, o outro também era atualizado.

Além disso, para garantir a comparabilidade entre diferentes estudos, foi necessário padronizar as unidades de concentração utilizadas nos ensaios de MIC, assegurando a uniformização dos valores. Esse ajuste foi realizado para que todas as medidas estivessem expressas na mesma unidade ($\mu\text{g/mL}$), permitindo análises estatísticas mais precisas e garantindo a integridade dos dados.

3.1.4 Integração e Refinamento dos Dados

Após a filtragem, os bancos de composição química e atividade antimicrobiana foram integrados, permitindo a correlação entre os perfis químicos dos OEs e seus respectivos efeitos antimicrobianos. A unificação foi realizada com base na correspondência entre os artigos presentes em ambas as bases, assegurando que cada amostra fosse corretamente associada aos seus respectivos resultados microbiológicos.

A partir dessa integração, foi possível estruturar um conjunto de dados adequado para futuras análises, permitindo a investigação da relação entre diferentes compostos químicos e seus efeitos antimicrobianos. O banco de dados consolidado fornece uma base robusta para análises estatísticas e a formulação de hipóteses sobre os efeitos biológicos das substâncias presentes nos OEs.

3.1.5 Curadoria e Pré-Processamento de Dados Químicos

O KNIME 5.1 (KNIME AG, Zurique, Suíça), uma plataforma de aprendizado de máquina de código aberto, foi utilizado para a integração, processamento, análise e exploração de dados (Beisken *et al.*, 2013; Ordenes; Silipo, 2021). Por meio de fluxos de trabalho personalizados, o KNIME facilitou a organização eficiente e a análise de dados, integrando nós de quimioinformática e algoritmos avançados de aprendizado de máquina. Essa integração possibilitou a implementação de análises sistemáticas e reproduzíveis, essenciais para a modelagem computacional e a avaliação biológica dos compostos estudados (Katchborian-Neto *et al.*, 2024; Ordenes; Silipo, 2021).

Durante o processo de curadoria, a primeira etapa envolveu o pré-processamento

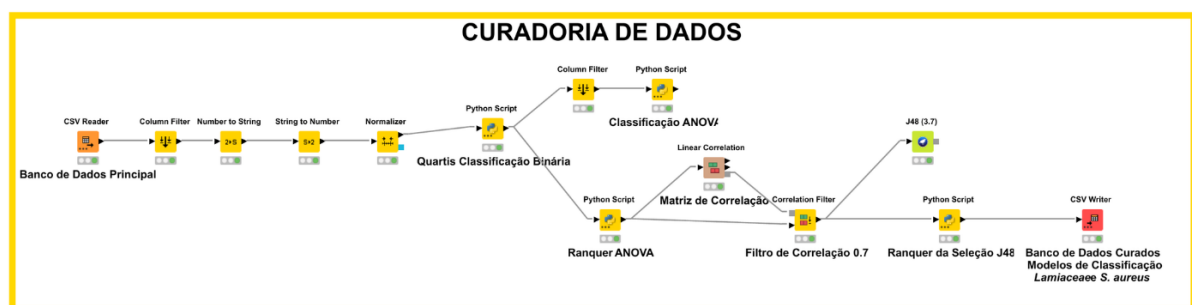
estatístico da coluna de dados de atividade no conjunto de dados principal. Devido à ampla variação nos valores de MIC, todas as medições foram inicialmente convertidas para $\mu\text{g/mL}$. Posteriormente, os valores de atividade passaram por processamento estatístico para aprimorar sua interpretabilidade. Inicialmente, os valores de MIC foram convertidos para uma escala logarítmica de base 10, com o objetivo de reduzir grandes variações e otimizar a distribuição dos dados. Após essa transformação, foi aplicada a técnica de autoscaling para normalizar a amplitude dos valores, permitindo comparações mais robustas e significativas dentro do conjunto de dados (Farias *et al.*, 2025). Por fim, os dados processados foram armazenados no formato .csv para futuras análises computacionais.

3.2 CONSTRUÇÃO DOS MODELOS DE PREDIÇÃO

3.2.1 Abordagem de Classificação

Foi desenvolvido um fluxo de trabalho na plataforma KNIME para classificar as variáveis mais correlacionadas com a atividade antibacteriana em cepas de *S. aureus*, conforme representado na Figura 1.

Figura 1 - Curadoria dos Dados e Geração dos Modelos de Predição



Fonte: Do autor (2025).

Inicialmente, o nó "Column Filter" foi utilizado para extrair colunas contendo informações relevantes para a construção do modelo de predição. Em seguida, foram aplicados nós de conversão de valores para garantir a correta interpretação dos valores numéricos pelos modelos. Para minimizar disparidades entre as variáveis na coluna de atividade, os dados auto escalonados foram normalizados em uma faixa de 0–1 usando o nó "Normalizer". Para conversão binária e agrupamento dos dados, necessário para a construção do modelo de classificação, foi implementado um script Python no nó "Python Script". Nesse estágio, foram

calculados os valores estatísticos médios da coluna de atividade processada e categorizados em duas classes: A (Ativo): MIC < 0.59 µg/mL e B (Inativo): MIC > 0.59 µg/mL.

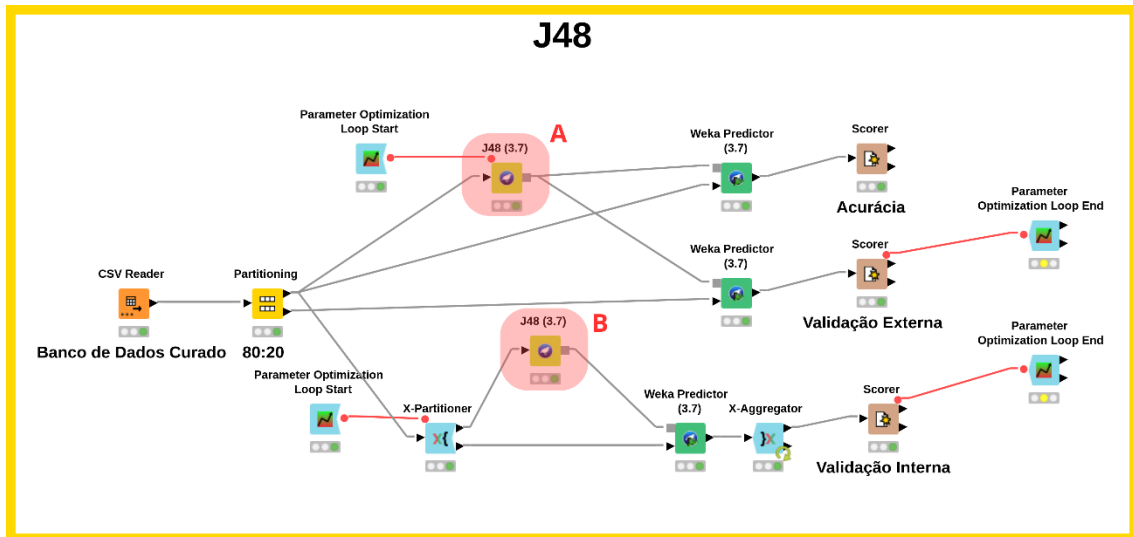
Posteriormente, um script Python foi empregado dentro do nó "Python Script" para realizar a classificação de variáveis usando Análise de Variância (ANOVA). Foram selecionadas variáveis com valores de $p \leq 0,2$ para a construção do modelo. Após essa seleção, foi aplicada a abordagem de Árvore de Decisão J48 (algoritmo de classificação baseado no C4.5) para classificar as variáveis mais correlacionadas com a atividade antibacteriana. Após a etapa de classificação, foi avaliado um limiar de correlação de similaridade >0,7 entre as variáveis selecionadas utilizando o nó "Correlation Filter". Nessa etapa, nenhuma variável foi excluída, pois todas apresentaram baixa correlação entre si. Esse processo garantiu a seleção das variáveis mais adequadas, otimizando sua similaridade e correlação com a atividade biológica.

3.2.2 Modelos de Predição

Os modelos de predição foram construídos utilizando as variáveis selecionadas (X) e os valores de MIC (Y) para avaliar sistematicamente e com precisão a atividade antibacteriana de diferentes compostos de OEs testados.

Para a avaliação do modelo de classificação, foram testados classificadores de árvore de decisão e redes neurais, como Redes Neurais Artificiais (ANN), NBTree (NBT), Random Forest (RF), Random Tree (RT) e J48, e seus resultados foram comparados usando métodos de aprendizado de máquina. Esta comparação utilizou um conjunto de dados de metabólitos caracterizados por CG-EM (Cromatografia Gasosa acoplada à Espectrometria de Massas) de OEs testados quanto à atividade contra cepas de *S. aureus* relatados pela literatura – é possível observar na Figura 2, o Fluxo de Trabalho desenvolvido na plataforma KNIME, sendo utilizado em todos os modelos, com troca apenas dos nós destacados como A e B (nós correspondentes aos algoritmos de classificação). O conjunto de dados foi particionado usando o nó "Particionamento" na plataforma KNIME, dividindo os dados em subconjuntos de treinamento e teste. Especificamente, 80% dos dados foram alocados para treinamento dos modelos, enquanto os 20% restantes serviram como um conjunto de testes externos. Este processo foi cuidadosamente concebido para validar o desempenho preditivo dos modelos e garantir a sua confiabilidade para aplicações práticas.

Figura 2 - Fluxo de trabalho utilizado na elaboração dos modelos de predição



Fonte: Do autor.

Legenda: as trocas dos nós A e B foram feitas para cada algoritmo de classificação - Random Forest (RF), Artificial Neural Networks (ANN), Random Tree (RT), Naïve Bayes Tree (NBT), e J48 Decision Tree.

3.2.3 Validação do Modelo

Ainda na plataforma KNIME, o nó "CSV Reader" foi usado para importar o conjunto de dados contendo informações de atividade biológica para cada composto e suas respectivas variáveis, servindo como base para o desenvolvimento dos modelos de predição. A validação interna foi realizada utilizando o nó "X-Partitioner", que dividiu aleatoriamente os dados em 10 grupos, enquanto os nós "X-Aggregator" e "Data to Report" foram empregados para consolidar e registrar os resultados. A eficácia do modelo foi avaliada usando métricas de desempenho padrão, incluindo precisão, que reflete a proporção de previsões corretas, sensibilidade, que mede a capacidade de identificar instâncias positivas e precisão, que quantifica a proporção de verdadeiros positivos entre todas as previsões positivas. O coeficiente de Cohen Kappa foi utilizado para medir a concordância entre as classificações previstas e as reais, levando-se em conta a concordância casual. Além disso, a F-measure, que integra precisão e recall, forneceu uma avaliação equilibrada do desempenho dos modelos. Valores elevados nessas métricas indicam um modelo de predição robusto e confiável, garantindo uma avaliação abrangente de suas capacidades preditivas dentro do conjunto de dados.

3.2.4 Ajuste de Hiperparâmetros

O ajuste de hiperparâmetros foi realizado para melhorar o desempenho dos modelos. O

processo de ajuste envolveu a varredura de parâmetros por meio de algoritmos de otimização de hiperparâmetros com validação cruzada, utilizando nós de "loop" e a estratégia de busca "Brute Force", onde todas as combinações de parâmetros possíveis foram avaliadas para identificar a configuração ideal dos modelos (Bagnall; Cawley, 2017; Bergstra *et al.*, 2011; Kurniawan; Ikhsan, 2020). Esse processo iterativo inicializa um loop de otimização de parâmetros, gerando vários conjuntos de parâmetros como variáveis de fluxo com base em uma estratégia de pesquisa predefinida. Esses parâmetros podem ser aplicados dentro do corpo do loop, diretamente ou convertendo-os em uma tabela de dados. Uma lista de parâmetros e seus respectivos intervalos de valores para o método da árvore de decisão J48 foram definidos e empregados.

4 RESULTADOS E DISCUSSÃO

4.1 COLETA DE DADOS E CONSTRUÇÃO DOS BANCOS DE DADOS

Conforme a metodologia, foram coletados cerca de 300 artigos científicos que abordam pesquisas a respeito da composição e atividade antimicrobiana de OEs de diversas plantas. Tais artigos foram selecionados com base em critérios pré-estabelecidos e mencionados na metodologia deste trabalho, necessitando fornecer informações a respeito da planta como: sua família, composição com no mínimo 70% dos constituintes identificados e atividade antimicrobiana frente a diversos microrganismos - porém o principal interesse foram as bactérias, o que não impossibilitou a coleta de dados para fungos.

Os dados de atividade de maior interesse foram os de MIC, por serem informações quantitativas, mas não se rejeitou os dados de Zona de Inibição, que apesar de serem dados qualitativos é possível utilizar-se dos padrões para estabelecer critérios de classificação da atividade dos óleos em questão. Os dados coletados foram utilizados na construção de um banco que aborda informações de atividade antimicrobiana (BDA) com cerca de 39 famílias de plantas e a partir destes foi desenvolvida um segundo banco que contém informações a respeito da composição (BDC) dos diversos óleos, como mais de 9000 entradas de atividade antimicrobiana para diferentes microrganismos.

O BDC foi desenvolvido separadamente do BDA (MIC e Zona de Inibição). Isso se mostrou necessário para facilitar a padronização dos nomes dos compostos descritos em cada artigo, já que esses nomes variam de autor para autor. Para realizar análises estatísticas, é essencial que instâncias idênticas sejam padronizadas, pois foi constatada uma variação significativa nas nomenclaturas utilizadas: enquanto alguns autores utilizam o nome usual, outros optam pelo nome IUPAC, por exemplo. Essas variações prejudicam o desenvolvimento de modelos de classificação e predição, uma vez que, para análises estatísticas precisas, é fundamental que instâncias idênticas possuam nomenclaturas uniformes, evitando que sejam erroneamente consideradas como distintas. Essa padronização de nomes foi feita e inclusa em uma coluna denominada IUPAC Name, e ao analisar os nomes IUPAC acrescentados ao BDC foi possível notar algumas variações e como garantia de diferenciação de instâncias, o SMILES Canônico se torna a melhor ferramenta de uniformização dos dados, já que é um dado que descreve exatamente qual substância foi relatada e é único para cada uma.

As características do BDC foram descritas e podem ser observadas no Quadro 1 a seguir:

Quadro 1 - Estrutura e descrição do banco de dados de composição

| Coluna | Descrição Detalhada do Banco de Dados de Composição |
|--------------------|--|
| Paper Number | Refere-se ao identificador único de cada artigo dentro do banco de dados, permitindo a rastreabilidade e organização das referências bibliográficas. |
| Author | Indica o primeiro autor do artigo, facilitando a identificação dos principais pesquisadores na área. |
| Year of Publishing | Especifica o ano de publicação do artigo, permitindo análises temporais e identificação de tendências na pesquisa. |
| Plant Number | Numera a posição em que os dados de cada planta aparecem no artigo, sendo crucial para artigos que descrevem múltiplas plantas. |
| Source | Fornece o nome científico da planta, essencial para a padronização e correta identificação das espécies estudadas. |
| Family | Identifica a família botânica da planta, contribuindo para análises filogenéticas e taxonômicas. |
| Composition | Lista os nomes das substâncias conforme relatados pelos autores, sendo fundamental para a comparação direta dos resultados entre diferentes estudos. |
| IUPAC Name | Padroniza os nomes das substâncias de acordo com a nomenclatura IUPAC, essencial para a uniformidade dos dados. |
| Canonical SMILES | Proporciona uma representação linear da estrutura química da substância, ignorando isomerias, o que é útil para diversas análises computacionais e uma maneira mais segura de uniformidade dos dados. |
| Isomeric SMILES | Detalha a estrutura química da substância, incluindo informações sobre sua isomeria, permitindo uma análise mais detalhada das propriedades químicas. Infelizmente nem todas as substâncias apresentaram tal dado. |
| Percentage | Informa a concentração em área de cada substância no óleo essencial, permitindo comparações quantitativas entre diferentes estudos. |

Fonte: Do autor (2025).

O BDA possui as primeiras colunas semelhantes ao BDC, com algumas outras informações a mais, sua descrição e estrutura podem ser observadas no Quadro 2 a seguir:

Quadro 2 - Estrutura e descrição do banco de dados de atividade

| Coluna | Descrição Detalhada do Banco de Dados de Atividade |
|-----------------------|---|
| Paper Number | Refere-se ao identificador único de cada artigo dentro do banco de dados, permitindo a rastreabilidade e organização das referências bibliográficas. |
| Author | Indica o primeiro autor do artigo, facilitando a identificação dos principais pesquisadores na área. |
| Year of Publishing | Especifica o ano de publicação do artigo, permitindo análises temporais e identificação de tendências na pesquisa. |
| Plant Number | Numera a posição em que os dados de cada planta aparecem no artigo, sendo crucial para artigos que descrevem múltiplas plantas. |
| Source | Fornece o nome científico da planta, essencial para a padronização e correta identificação das espécies estudadas. |
| Family | Identifica a família botânica da planta, contribuindo para análises filogenéticas e taxonômicas. |
| Percentage Identified | Indica o total da concentração em área identificada do óleo essencial em questão. |
| Method | Indica se a atividade biológica está descrita em MIC ou Zona de Inibição. |
| Microorganism | Indica o microrganismo específico para o qual o óleo essencial foi testado. Foram coletados dados com foco em bactérias, mas também há dados para fungos. |
| Biological Activity | Indica os valores da atividade biológica avaliada contra determinado microrganismo. |

Fonte: Do autor (2025).

Com ambos os bancos de dados construídos e padronizados, foi necessário avaliar os dados e discutir as abordagens a serem seguidas para o desenvolvimento dos métodos computacionais e a elaboração dos modelos de predição de atividade. Inicialmente, pretendia-se construir modelos com um domínio de aplicação muito abrangente (para diversas plantas e diversos microrganismos distintos), o que, na prática, seria dificultado pelo volume de dados em cada banco (>9000 entradas no BDA e >18000 entradas no BDC) e pelo fato de a variabilidade dos dados ser alta, sendo necessário diminuí-la. Portanto, optou-se por desenvolver modelos ainda abrangentes em relação às diferentes plantas e famílias, mas específicos para o microrganismo *S. aureus* e para a família Lamiaceae, considerando dados apenas de MIC (massa/volume – m/v), padronizados para unidade de $\mu\text{g/mL}$. Essa escolha foi feita levando em consideração o BDA e considerando os microrganismos que mais geraram dados possíveis de serem correlacionados (considerando OEs com mais de 70% de sua composição total identificada, método de MIC (m/v) com valores acima de zero, decimais ou

inteiros, sem intervalos e sem símbolos de $>$, \geq , $<$, \leq , e desconsiderando células vazias ou sem valores) foram *Candida albicans* (141 entradas de dados), *P. aeruginosa* (156 entradas de dados) *E. coli* (282 entradas de dados), e a selecionada *S. aureus* (310 entradas de dados).

4.2 BANCO DE DADOS PRINCIPAL PARA MODELO DE PREDIÇÃO DE ATIVIDADE DE ÓLEOS ESSENCIAIS DE ESPÉCIES DA FAMÍLIA LAMIACEAE CONTRA *Staphylococcus aureus*

O conjunto de dados para a construção dos modelos de predição foi compilado a partir de dados da literatura – a coleção de artigos coletados encontra-se listada no Apêndice A (cerca de 34 artigos que mencionam plantas da família Lamiaceae com atividade relatada para *S. aureus*). Para a construção do banco de dados final, um total de 117 OEs e 390 compostos obtidos de diferentes partes de plantas da família Lamiaceae com atividade relatada contra cepas de *S. aureus* foram selecionados com base em sua diversidade química. As propriedades antibacterianas descritas para esses compostos serviram como base para o desenvolvimento dos modelos de predição neste estudo.

4.2.1 Seleção de atributos e critérios de desempenho para elaboração dos modelos a partir do Banco de Dados Principal

Para avaliar a correlação entre os compostos dos OEs (Variável X) e seus valores de MIC (Variável Y), os 390 compostos foram submetidos a processamento estatístico usando um fluxo de trabalho curado na plataforma KNIME. Foi aplicado um filtro de Análise de Variância (ANOVA), selecionando 214 variáveis com valores de $p \leq 0,2$. Em seguida, foram removidos 149 compostos altamente correlacionados (correlação $>0,7$), refinando o conjunto de dados para 64 variáveis. Por fim, foram identificados 5 compostos chave com base na sua correlação com a atividade biológica utilizando o algoritmo de árvore de decisão J48. Esses compostos selecionados foram posteriormente usados como recursos de entrada para desenvolver alguns modelos de aprendizado de máquina, incluindo ANN, RF, NBT, RT e J48. Essas entradas de dados foram separadas em 80% dos dados para treino e 20% para teste e foram listadas para melhor visualização na Tabela 1 e Tabela 2, respectivamente a seguir:

Tabela 1 - Dados de treino curados - utilizados para elaboração dos modelos de predição
(continua)

| M203 <chem>CC1=CCCC(=C)C2 CC(C2CC1)(C)C</chem> | M103 <chem>CC1(C2CCC(O1) (CC2)C)C</chem> | M100 <chem>CC1(C2CCC(C2)C1=C)C</chem> | M97 <chem>CC1(C2CCC(C2)(C1=O)C)C</chem> | M241 <chem>CC1CCC(C2C13C2 C(=CC3)C)C(C)C</chem> | Classe |
|---|---|--|--|--|--------|
| 2.9 | 0.1 | 0 | 0 | 0 | B |
| 8.2 | 22 | 0 | 0 | 0 | B |
| 4.1 | 16.3 | 0 | 0 | 0 | B |
| 2.6 | 17.5 | 0 | 0 | 0 | B |
| 0.12 | 2.47 | 2.75 | 0 | 0 | A |
| 0.22 | 10.04 | 8.73 | 0 | 0 | A |
| 2.28 | 0.4 | 1.08 | 0 | 0 | A |
| 0 | 14.12 | 1.49 | 0 | 0 | A |
| 0.85 | 19.96 | 1.83 | 0 | 0 | A |
| 7.05 | 0 | 0 | 0 | 0.17 | A |
| 2.53 | 3.1 | 0.54 | 0 | 0.82 | A |
| 1.34 | 54.3 | 0 | 0 | 0 | A |
| 0.52 | 0 | 5.91 | 0 | 0 | A |
| 5.13 | 0.19 | 0 | 0 | 0.08 | A |
| 0 | 0.001 | 0.001 | 0 | 0.08 | B |
| 0.9 | 14.3 | 0 | 0 | 0 | B |
| 0.4 | 4.5 | 0.1 | 0.3 | 0 | B |
| 0.2 | 3.2 | 0.1 | 0.2 | 0 | B |
| 0.3 | 9.3 | 0.1 | 0.2 | 0 | B |
| 0.4 | 3.4 | 0.1 | 0.3 | 0 | B |
| 11 | 0.1 | 0 | 1 | 0 | B |
| 1.6 | 9.2 | 0.1 | 1.2 | 0 | B |
| 1.2 | 5.9 | 0 | 0.3 | 0 | B |
| 1.4 | 10.2 | 0 | 0.5 | 0 | B |
| 1.16 | 25.52 | 5.59 | 0.001 | 0 | B |
| 6.18 | 9.69 | 7.09 | 0.001 | 0 | B |
| 2.7 | 7.45 | 3.45 | 0.2 | 0 | B |
| 1.31 | 0.001 | 0 | 0 | 0 | B |

Tabela 1 - Dados de treino curados - utilizados para elaboração dos modelos de predição
(continuação)

| M203 <chem>CC1=CCCC(=C)C2 CC(C2CC1)(C)C</chem> | M103 <chem>CC1(C2CCC(O1) (CC2)C)C</chem> | M100 <chem>CC1(C2CCC(C2)C1=C)C</chem> | M97 <chem>CC1(C2CCC(C2)(C1=O)C)C</chem> | M241 <chem>CC1CCC(C2C13C2 C(=CC3)C)C(C)C</chem> | Classe |
|---|---|--|--|--|--------|
| 0.71 | 0.15 | 0 | 0 | 0 | B |
| 2.53 | 0 | 0 | 0 | 0 | B |
| 2.43 | 0 | 0 | 0 | 0 | B |
| 2.45 | 0 | 0 | 0 | 0 | B |
| 2.98 | 0 | 0 | 0 | 0 | B |
| 0 | 0 | 1.2 | 0 | 0 | A |
| 0 | 0 | 0 | 0 | 0 | B |
| 1.11 | 10.56 | 3.13 | 0 | 0 | A |
| 0.9 | 1.2 | 0.2 | 0 | 0 | B |
| 1.62 | 0 | 0.44 | 0 | 0 | B |
| 1.09 | 0 | 0 | 0 | 0 | B |
| 3 | 0 | 0 | 0 | 0 | B |
| 2.3 | 1.1 | 2.4 | 0 | 0.5 | A |
| 0.6 | 27.9 | 2.6 | 0 | 0 | A |
| 0.2 | 14.9 | 0 | 0 | 0 | A |
| 1 | 15.5 | 2.8 | 0 | 0 | A |
| 0.2 | 61.4 | 3.4 | 0 | 0 | A |
| 0.2 | 61.4 | 3.4 | 0 | 0 | A |
| 0.8 | 50.7 | 3.4 | 0 | 0 | A |
| 0.6 | 60.9 | 2.9 | 0 | 0 | A |
| 0.1 | 12.7 | 0.2 | 0 | 0 | A |
| 0.8 | 50.7 | 3.4 | 0 | 0 | A |
| 0.6 | 60.9 | 2.9 | 0 | 0 | A |
| 0 | 0 | 0 | 0 | 0 | B |
| 4.22 | 0 | 0 | 0 | 0 | B |
| 2.53 | 0 | 0 | 0 | 0 | B |
| 3.55 | 0.9 | 0.16 | 0 | 0 | B |
| 2.55 | 1.06 | 0 | 0 | 0 | B |

Tabela 1 - Dados de treino curados - utilizados para elaboração dos modelos de predição
(continuação)

| M203 CC1=CCCC(=C)C2 CC(C2CC1)(C)C | M103 CC1(C2CCC(O1)(CC2)C)C | M100 CC1(C2CCC(C2)C1=C)C | M97 CC1(C2CCC(C2)(C1=O)C)C | M241 CC1CCC(C2C13C2C(=CC3)C)C(C)C | Classe |
|---|-------------------------------|-----------------------------|-------------------------------|--------------------------------------|--------|
| 2.02 | 0 | 0 | 0 | 0 | B |
| 0.95 | 1.66 | 0.57 | 0 | 0 | A |
| 2.31 | 0 | 0 | 0 | 0 | A |
| 0 | 0 | 0 | 0 | 0 | A |
| 1.08 | 0 | 0 | 0 | 0 | A |
| 2.66 | 0 | 0 | 0 | 0 | A |
| 2.62 | 0 | 0 | 0 | 0 | A |
| 2.1 | 0 | 0 | 0 | 0 | B |
| 11.8 | 15.7 | 1.85 | 0 | 0 | B |
| 11.8 | 15.7 | 1.85 | 0 | 0 | B |
| 4.54 | 3.02 | 6.54 | 0 | 0 | B |
| 0.73 | 0 | 0.3 | 0 | 0.1 | A |
| 1.16 | 3.11 | 0.21 | 0 | 0 | B |
| 1.16 | 3.11 | 0.21 | 0 | 0 | B |
| 0.78 | 0 | 0.02 | 0 | 0 | A |
| 3.62 | 0 | 0 | 0 | 0 | A |
| 3.62 | 0 | 0 | 0 | 0 | A |
| 3.15 | 29.2 | 5.99 | 0 | 0 | B |
| 0.15 | 29.2 | 5.99 | 0 | 0 | B |
| 2.75 | 27.4 | 2.35 | 0 | 0 | B |
| 1 | 5.65 | 0.5 | 0 | 0 | B |
| 1 | 5.65 | 0.5 | 0 | 0 | B |
| 13.4 | 0 | 0 | 0 | 0 | B |
| 9 | 0 | 0 | 0 | 0 | B |
| 0 | 0 | 0 | 0 | 0 | B |
| 12.7 | 0 | 0 | 0 | 0 | B |
| 0 | 4.7 | 0 | 0 | 0 | A |
| 0 | 42.9 | 2.57 | 0 | 0 | A |
| 0.88 | 0.4 | 0 | 0 | 0 | A |

Tabela 1 - Dados de treino curados - utilizados para elaboração dos modelos de predição (conclusão)

| M203 CC1=CCCC(=C)C2 CC(C2CC1)(C)C | M103 CC1(C2CCC(O1)(CC2)C)C | M100 CC1(C2CCC(C2)C1=C)C | M97 CC1(C2CCC(C2)(C1=O)C)C | M241 CC1CCC(C2C13C2C(=CC3)C)C(C)C | Classe |
|---|-------------------------------|-----------------------------|-------------------------------|--------------------------------------|--------|
| 4.01 | 47.9 | 0 | 0 | 0 | A |
| 2.06 | 0.98 | 0 | 0 | 0 | A |
| 1.62 | 8.13 | 2.09 | 0 | 0 | A |
| 1.37 | 8.07 | 4.32 | 0 | 0 | A |
| 1.37 | 7.54 | 3.55 | 0 | 0 | A |
| 2.03 | 0.4 | 0 | 0 | 0 | A |
| 2.13 | 0.93 | 0 | 0 | 0 | A |
| 1.17 | 1.5 | 0 | 0 | 0 | A |
| 2 | 51 | 6.3 | 0 | 0 | A |

Fonte: Do autor (2025).

Legenda: A = ativo;
B = inativo.

Tabela 2 - Dados de teste curados - utilizados para elaboração dos modelos de predição (continua)

| M203 CC1=CCCC(=C)C2 CC(C2CC1)(C)C | M103 CC1(C2CCC(O1)(CC2)C)C | M100 CC1(C2CCC(C2)C1=C)C | M97 CC1(C2CCC(C2)(C1=O)C)C | M241 CC1CCC(C2C13C2C(=CC3)C)C(C)C | Classe |
|---|-------------------------------|-----------------------------|-------------------------------|--------------------------------------|--------|
| 1.78 | 0.98 | 0.02 | 0 | 0 | A |
| 0 | 3.45 | 0 | 0 | 0 | A |
| 1.12 | 14.3 | 5.36 | 0 | 0 | A |
| 0.3 | 1.8 | 0 | 0.2 | 0 | B |
| 0.3 | 6.4 | 0 | 0.2 | 0 | B |
| 0.3 | 2 | 0.1 | 0.1 | 0 | B |
| 0.09 | 5.45 | 17.19 | 0 | 0 | A |
| 2.36 | 0 | 0 | 0 | 0 | B |
| 0.17 | 11.91 | 5.7 | 0 | 0 | A |
| 2.5 | 0 | 0 | 0 | 0 | B |
| 0.3 | 34.3 | 1.6 | 0 | 0 | A |
| 0.3 | 58.1 | 4.2 | 0 | 0 | A |
| 0 | 0 | 0 | 0 | 0 | B |
| 2.75 | 6.36 | 0 | 0 | 0 | B |
| 2 | 1.05 | 0 | 0.53 | 0 | A |
| 0.24 | 31.99 | 17.06 | 0 | 0.17 | A |
| 0 | 0 | 0 | 0 | 0.76 | A |
| 0.78 | 0 | 0.02 | 0 | 0 | B |
| 2.75 | 27.4 | 2.35 | 0 | 0 | A |
| 4.5 | 0 | 0 | 0 | 0 | B |
| 1.84 | 0 | 0 | 0 | 0 | A |

Tabela 2 - Dados de teste curados - utilizados para elaboração dos modelos de predição (conclusão)

| M203 | M103 | M100 | M97 | M241 | Classe |
|--|------------------------------------|----------------------------------|-------------------------------------|---|--------|
| <chem>CC1=CCCC(=C)C2CC(C2CC1)(C)C</chem> | <chem>CC1(C2CCC(O1)(CC2)C)C</chem> | <chem>CC1(C2CCC(C2)C1=C)C</chem> | <chem>CC1(C2CCC(C2)(C1=O)C)C</chem> | <chem>CC1CCC(C2C13C2C(=CC3)C)C(C)C</chem> | |
| 0 | 0.5 | 0.1 | 0 | 0 | A |
| 4 | 9.6 | 4 | 0 | 0 | A |
| 0.44 | 2.51 | 0 | 0 | 0 | A |

Fonte: Do autor (2025).

Legenda: A = ativo;

B = inativo.

4.2.2 Construção dos Modelos de Predição de Atividade Antibacteriana contra *Staphylococcus aureus*

Utilizando os 5 compostos selecionados de 117 OEs derivados de plantas da família Lamiaceae, foram desenvolvidos modelos de predição baseados em Redes Neurais Artificiais (ANNs) e Classificadores de Árvore de Decisão (Tabela 3).

Tabela 3 - Métricas de desempenho para modelos

(continua)

| Algoritmo de Machine Learning | Métricas | Acurácia (%) | Cohen's Kappa | Sensibilidade | Especificidade | Precisão | Recall | F-measure |
|-------------------------------|-------------------|--------------|---------------|---------------|----------------|----------|--------|-----------|
| ANN | Acurácia | 74.10 | 0.472 | 0.86 | 0.605 | 0.793 | 0.860 | 0.779 |
| | Validação Externa | 72.2 | 0.464 | 0.938 | 0.550 | 0.625 | 0.938 | 0.750 |
| | Validação Interna | 66.7 | 0.309 | 0.93 | 0.368 | 0.625 | 0.930 | 0.748 |
| *J48 | Acurácia | 80.6 | 0.613 | 0.818 | 0.796 | 0.783 | 0.818 | 0.800 |
| | Validação Externa | 79.2 | 0.565 | 0.800 | 0.778 | 0.857 | 0.800 | 0.828 |
| | Validação Interna | 79.6 | 0.591 | 0.795 | 0.796 | 0.778 | 0.795 | 0.787 |
| RF | Acurácia | 94.6 | 0.892 | 0.932 | 0.959 | 0.953 | 0.932 | 0.943 |
| | Validação Externa | 58.3 | 0.184 | 0.533 | 0.667 | 0.727 | 0.533 | 0.615 |
| | Validação Interna | 69.9 | 0.398 | 0.705 | 0.694 | 0.674 | 0.705 | 0.689 |
| RT | Acurácia | 96.8 | 0.935 | 0.932 | 1.000 | 1.000 | 0.932 | 0.965 |
| | Validação Externa | 58.3 | 0.216 | 0.467 | 0.778 | 0.778 | 0.467 | 0.583 |
| | Validação Interna | 66.7 | 0.337 | 0.727 | 0.612 | 0.627 | 0.727 | 0.674 |

Tabela 3 - Métricas de desempenho para modelos

(conclusão)

| Algoritmo de Machine Learning | Métricas | Acurácia (%) | Cohen's Kappa | Sensibilidade | Especificidade | Precisão | Recall | F-measure |
|-------------------------------|-------------------|--------------|---------------|---------------|----------------|----------|--------|-----------|
| NBT | Acurácia | 81.7 | 0.633 | 0.795 | 0.837 | 0.814 | 0.795 | 0.805 |
| | Validação Externa | 75.0 | 0.489 | 0.733 | 0.778 | 0.846 | 0.733 | 0.786 |
| | Validação Interna | 65.6 | 0.318 | 0.750 | 0.571 | 0.611 | 0.750 | 0.673 |

Fonte: Do autor (2025).

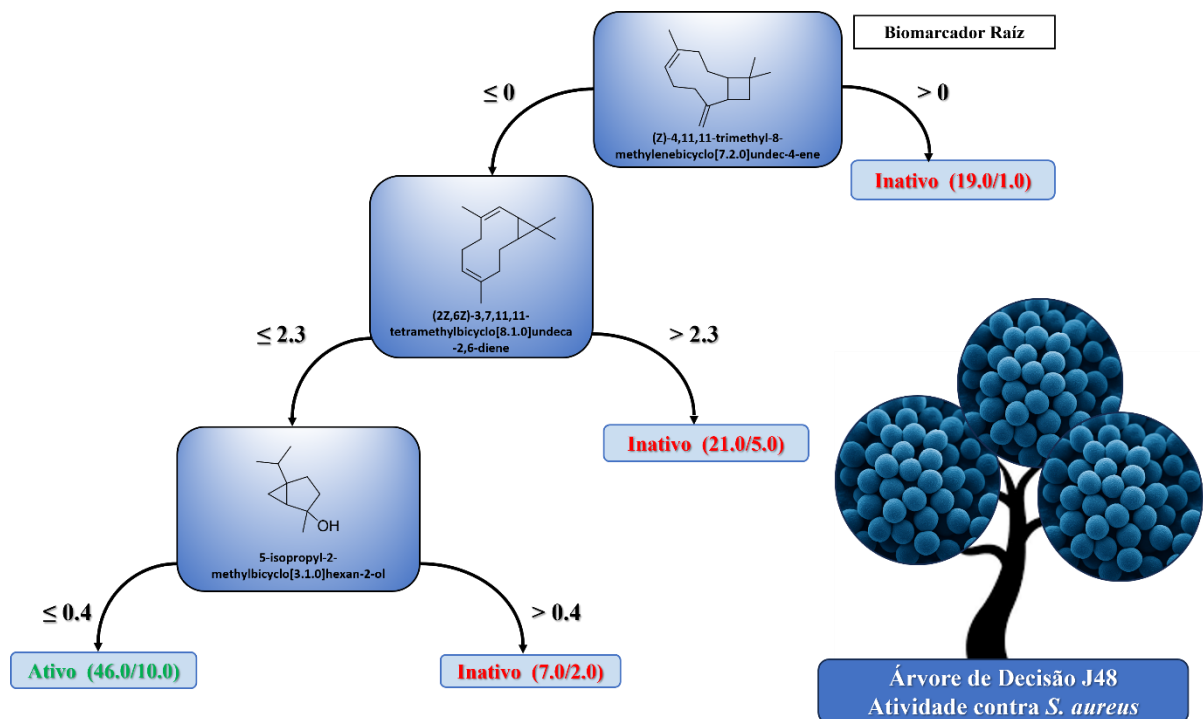
Legenda: *Melhor Modelo por comparação entre Acurácia, Kappa de Cohen, Sensibilidade, Especificidade, Precisão, Recall e F-measure.

Para avaliar a qualidade preditiva dos modelos gerados, foram consideradas três principais métricas de desempenho: acurácia, validação interna e validação externa (John *et al.*, 2021; Khan; Roy, 2021). A precisão é uma métrica amplamente utilizada que avalia a capacidade geral do modelo de classificar corretamente as instâncias em suas respectivas categorias. No entanto, em casos de desequilíbrio de classes, é crucial considerar métricas complementares. A validação interna avalia o desempenho do modelo usando o conjunto de dados de treinamento, normalmente empregando técnicas de validação cruzada para estimar sua generalização para dados desconhecidos. Igualmente importante, a validação externa serve como um teste crítico da capacidade de generalização dos modelos para dados novos e inéditos que não foram usados durante o treinamento ou validação interna. Este conjunto de testes independente é essencial para demonstrar a confiabilidade do modelo em fazer previsões precisas diante de dados nunca vistos antes (Zhan; Yu, 2020; Zhang *et al.*, 2006). Para validar o modelo, o conjunto de dados deve ser dividido em um conjunto de treinamento e um conjunto de testes, sendo este último usado para avaliar o desempenho preditivo. É crucial que ambos os conjuntos englobem um amplo espectro de atividades biológicas e diversidade estrutural para garantir previsões robustas e generalizáveis (PINGAEW *et al.*, 2022; VAIDYA *et al.*, 2014).

Para que um modelo de predição seja considerado satisfatório, este deve apresentar acurácia e valores de validação externa superiores a 70% (Balim *et al.*, 2017; Ion *et al.*, 2019; Jamshidnezhad *et al.*, 2021). Com base nos resultados (Tabela 3), o modelo Random Forest (RF) demonstrou a maior precisão (94,6%) e o coeficiente Kappa de Cohen (0,892) na validação interna, sugerindo fortes capacidades preditivas dentro do conjunto de dados de treinamento. Da mesma forma, o modelo Random Tree (RT) também apresentou alta precisão (96,8%) e Cohen Kappa (0,935), indicando seu potencial em classificar a atividade antibacteriana de forma eficaz. No entanto, tanto o RF como o RT apresentaram um fraco desempenho de

validação externa (58,3% de acurácia e Cohen Kappa < 0,25), sugerindo possível sobreajuste, onde os modelos apresentam um desempenho excepcionalmente bom nos dados de treino, mas não conseguem generalizar para dados inéditos que não foram usados durante o treinamento do modelo. Entre todos os modelos, o J48 Decision Tree – representado pela árvore de decisão na Figura 3 - surgiu como o algoritmo de melhor desempenho em termos de precisão equilibrada em treinamento, validação interna e validação externa. Ele alcançou uma acurácia de 80,6% no treinamento, 79,6% na validação interna e 79,2% na validação externa, com Cohen Kappa consistentemente acima de 0,55 em todas as validações. Isso indica que o J48 fornece um equilíbrio satisfatório entre precisão e generalização, tornando-o dentre os modelos gerados, o mais confiável para prever a atividade antibacteriana contra *S. aureus* a partir do uso de OEs de plantas da família Lamiaceae.

Figura 3 - Árvore de Decisão J48 - Modelo de Predição para Lamiaceae contra *S. aureus*



Fonte: Do autor (2025).

O algoritmo J48 Decision Tree, um derivado do algoritmo C4.5, é amplamente reconhecido por sua capacidade de gerar modelos de classificação interpretáveis com base em regras de decisão hierárquicas. Estudos têm destacado que as árvores de decisão têm bom desempenho em tarefas de previsão de bioatividade, especialmente quando se trata de conjuntos

de dados contendo diversas estruturas químicas e atividades biológicas (Kaur; Chhabra, 2014). O desempenho do modelo J48 Decision Tree foi ainda mais otimizado através do ajuste de hiperparâmetros, garantindo um equilíbrio entre precisão, generalização e interpretabilidade. Durante o processo de hiperparametrização, diferentes combinações de parâmetros foram testadas usando estratégias de busca sistemática, como *Brute Force Search*, para identificar a configuração ideal. Esse ajuste fino foi crucial para melhorar o desempenho do modelo, reduzir possíveis vieses e melhorar sua capacidade de generalização frente a novos dados (Tabela 4).

Tabela 4 - Ajuste de hiperparâmetros ótimos do modelo de predição J48

| Modelo | Parâmetros | Range | Selection Value |
|--------|--------------------|------------|-----------------|
| J48 | C_confidenceFactor | 0.001-1.00 | 0.04 |
| | M_minNumObj | 1-200 | 6 |
| | N_numfolds | 1-200 | 3 |
| | S_seed | 1-5000 | 1 |
| | Unpruned | True-False | False |
| | useLaplace | True-False | False |
| | useMDLcorrection | True-False | True |
| | subtreeRasing | True-False | True |

Fonte: Do autor (2025).

O modelo de árvore de decisão J48 hiperparametrizado apresentou quatro folhas e sete nós internos, indicando uma estrutura relativamente simples e de fácil interpretação. Foram utilizadas pelo modelo três moléculas como atributos principais para a classificação dos OEs em ativos (A) e inativos (B) contra *S. aureus*.

A substância (Z)-4,11,11-trimethyl-8-methylenebicyclo[7.2.0]undec-4-ene (SMILES: C/C/1=C/CCC(=C)C2CC(C2CC1)(C)C) - quando não presente (concentração ≤ 0) no óleo essencial, classifica as amostras de óleo como inativa (B) com alta precisão (19 acertos e apenas 1 erro). Isso sugere que essa molécula pode desempenhar um papel central na atividade antimicrobiana dos OEs testados. Já a substância (2Z,6Z)-3,7,11,11-tetramethylbicyclo[8.1.0]undeca-2,6-diene.

(SMILES: C/C/1=C/CC/C(=C\C2C(C2(C)C)CC1)/C) - se presente em alta concentração (>2.3), o óleo essencial é classificado como inativo (B) em 21 amostras (com 5 erros). Esse resultado sugere que essa molécula pode estar associada à redução da atividade antimicrobiana, possivelmente por interferência na ação dos compostos bioativos. Entretanto, a substância 5-isopropyl-2-methylbicyclo[3.1.0]hexan-2-ol (SMILES: CC(C)C12CCC(C1C2)(C)O) - se

presente em baixa concentração (≤ 0.4), classifica a amostra como Ativa (A) com 46 acertos e 10 erros e quando sua concentração for superior a 0.4, a amostra é classificada como inativa (B) (7 amostras, com 2 erros). Isso indica que essa molécula pode desempenhar um papel sinérgico na atividade antimicrobiana quando presente em concentrações moderadas, mas pode exercer efeitos antagonistas em altas concentrações (Zago *et al.*, 2009; Ferreira *et al.*, 2023).

5 CONCLUSÃO

A resistência bacteriana a antibióticos representa um desafio crescente à saúde pública, exigindo alternativas terapêuticas inovadoras e eficazes. Neste contexto, os OEs surgem como uma promissora fonte de compostos bioativos com potencial antibacteriano. No entanto, a identificação de OEs eficazes e a compreensão dos mecanismos que determinam sua bioatividade são desafios que exigem abordagens sistemáticas e precisas. Diante desse cenário, este estudo desenvolveu uma abordagem computacional para predição da atividade antibacteriana de OEs da família Lamiaceae, combinando quimioinformática, aprendizado de máquina e estatística multivariada. A partir de uma ampla revisão bibliográfica e da construção de um banco de dados estruturado contendo informações sobre a composição química e a atividade antimicrobiana dos OEs, foi possível estabelecer correlações significativas entre os constituintes químicos e sua eficácia contra *S. aureus*.

A aplicação de algoritmos de aprendizado de máquina permitiu a identificação de padrões químicos associados à bioatividade e possibilitou a construção de modelos de predição confiáveis de acordo com as métricas obtidas. Dentre os algoritmos testados, a Árvore de Decisão J48 se destacou como o modelo mais eficiente, apresentando 80,6% de acurácia na validação externa, o que reforça o potencial da modelagem computacional na triagem de compostos antimicrobianos. Além de fornecer uma metodologia diferenciada para análise e predição da bioatividade de OEs, este estudo demonstra a importância da integração entre química e inteligência artificial para otimizar o processo de descoberta de novos agentes terapêuticos. Os modelos desenvolvidos podem ser utilizados como ferramentas para guiar futuras investigações experimentais, reduzindo o tempo e os custos associados à triagem de produtos naturais.

Os resultados obtidos abrem perspectivas para a aplicação dessas abordagens computacionais não apenas na pesquisa de agentes antibacterianos, mas também na descoberta de compostos bioativos com outras propriedades biológicas, como atividades antifúngica, antioxidante e anti-inflamatória. Assim, este estudo não apenas reafirma o potencial dos OEs como alternativas aos antibióticos convencionais, mas também apresenta um modelo de predição computacional que pode ser expandido e aprimorado para diversas aplicações biomédicas e farmacêuticas. A continuidade dessa pesquisa poderá incluir a validação experimental dos compostos preditos a partir dos testes com OEs coletados pelo grupo de pesquisa, bem como o desenvolvimento de modelos para diferentes espécies vegetais e patógenos que não foram abordados neste estudo inicialmente, consolidando essa abordagem

como uma ferramenta promissora na investigação e combate a doenças causadas por patógenos em geral.

REFERÊNCIAS

- ABERS, Mareshah *et al.* Antimicrobial activity of the volatile substances from essential oils. **BMC Complementary Medicine And Therapies**, [s.l.], v. 21, n. 1, p. 1-14, Apr. 2021.
- AL-KURAIISHI, Nawar *et al.* Neuroprotective potential of *Rosmarinus officinalis* L.: Mechanisms of action and therapeutic prospects. **Applied Sciences**, [s.l.], v. 14, n. 15, art. 6417, July 2024.
- ALI, Muhammad *et al.* Exploring Lamiaceae-derived bioactive compounds as nature's arsenal for sustainable pest management. **Phytochemistry Reviews**, [s.l.], v. 24, n. 1, p. 1-26, June 2024.
- ALVES, Vinicius *et al.* Quimioinformática: uma introdução. **Química Nova**, [s.l.], v. 41, n. 2, p. 202-212, 08 nov. 2017.
- BAGNALL, Anthony J.; CAWLEY, Gavin C. On the use of default parameter settings in the empirical evaluation of classification algorithms. **CoRR - Computing Research Repository**, [s.l.], v. abs/1703.06777, Sept. 2017.
- BALDIM, João Luiz *et al.* The Correlation between Chemical Structures and Antioxidant, Prooxidant, and Antitrypanosomatid Properties of Flavonoids. **Oxidative Medicine And Cellular Longevity**, [s.l.], v. 2017, n. 1, p. 1-12, June 2017.
- BALDIM, João Luiz *et al.* The synergistic effects of volatile constituents of *Ocimum basilicum* against foodborne pathogens. **Industrial Crops And Products**, [s.l.], v. 112, p. 821-829, Feb. 2018.
- BASSOLÉ, I. H. N.; JULIANI, H. R. Essential oils in combination and their antimicrobial properties. **Molecules**, [s.l.], v. 17, n. 4, p. 3989–4006, Apr. 2012.
- BEISKEN, Stephan *et al.* KNIME-CDK: workflow-driven cheminformatics. **BMC Bioinformatics**, [s.l.], v. 14, n. 1, p. 1-4, Aug. 2013.
- BERGSTRA, James *et al.* Algorithms for Hyper-Parameter Optimization. **25th Annual Conference On Neural Information Processing Systems**, [s.l.], v. 1, p. 1-9, Nov. 2011.
- BEJENARU, Ludovic Everard *et al.* Polyphenols Investigation and Antioxidant and Anticholinesterase Activities of *Rosmarinus officinalis* L. Species from Southwest Romania Flora. **Molecules**, [s.l.], v. 29, n. 18, p. 1-20, Sept. 2024.
- BRANDT, Christian *et al.* The bigger picture: the history of antibiotics and antimicrobial resistance displayed by scientometric data. **International Journal Of Antimicrobial Agents**, [s.l.], v. 44, n. 5, p. 424-430, Nov. 2014.
- CANUTO, Gisele *et al.* Metabolômica: definições, estado-da-arte e aplicações representativas. **Química Nova**, [s.l.], v. 41, n. 1, p. 75-91, out. 2017.

CARVALHO, Zuleide Silva de. Potencial fitoquímico dos óleos essenciais: exploração e aplicações. **Boletim Científico Agrônomo do CCAAB/UFRB**, [s.l.], v. 1, p. 1-9, set. 2023.

CHAGAS-PAULA, Daniela *et al.* A Metabolomic Approach to Target Compounds from the Asteraceae Family for Dual COX and LOX Inhibition. **Metabolites**, [s.l.], v. 5, n. 3, p. 404-430, July 2015.

CHIEN, Li-Hsiang *et al.* Effects of lavender aromatherapy on sleep quality and physiological indicators in patients with ischemic heart disease: A randomized controlled trial. **Evidence-Based Complementary and Alternative Medicine**, [s.l.], v. 2023, art. 10671255, 2023.

D, Archi; K, Pooja; F, Shital. Formulation and Evaluation of Calendula and Lavender Oil Based Topical Antifungal Gel. **Journal Of Pharmaceutical Research International**, [s.l.], v. 36, n. 7, p. 133-147, 17 June 2024.

DHANE, Machhindranath Manikrao *et al.* Classification And Prediction Of Essential Oils Using Mobile Nets. **Natural Volatiles & Essential Oils**, [s.l.], v. 8, n. 6, p. 317-325, Jan. 2022.

DIÉGUEZ-SANTANA, Karel; GONZÁLEZ-DÍAZ, Humberto. *Machine learning* in antibacterial discovery and development: a bibliometric and network analysis of research hotspots and trends. **Computers In Biology And Medicine**, [s.l.], v. 155, p. 106638, Feb. 2023.

DUNN, Warwick B.; ELLIS, David. I.. Metabolomics: current analytical platforms and methodologies. **Trac Trends In Analytical Chemistry**, [s.l.], v. 24, n. 4, p. 285-294, Apr. 2005.

EBADOLLAHI, Asgar; ZIAEE, Masumeh; PALLA, Franco. Essential Oils Extracted from Different Species of the Lamiaceae Plant Family as Prospective Bioagents against Several Detrimental Pests. **Molecules**, [s.l.], v. 25, n. 7, p. 1-15, 28 Mar. 2020.

ELBOUZIDI, Amine *et al.* Formulation of a Three-Component Essential Oil Mixture from *Lavandula dentata*, *Rosmarinus officinalis*, and *Myrtus communis* for Improved Antioxidant Activity. **Pharmaceuticals**, [s.l.], v. 17, n. 8, p. 1-21, Aug. 2024.

FARIAS, Bárbara F. *et al.* Computational Modeling and Biological Evaluation of Benzophenone Derivatives as Antileishmanial Agents. **Journal Of The Brazilian Chemical Society**, [s.l.], v. 36, n. 8, p. 1-8, 2025.

FARRAH, S. R. *et al.* Essential oils for biofilm control: mechanisms, synergies, and applications. **Antibiotics**, [s.l.], v. 14, n. 5, p. 503, 2025.

FERREIRA, Nicolly Soares *et al.* Sinergismo antimicrobiano e atividades antibiofilme dos óleos essenciais de *Mentha piperita* L. e *Eucalyptus globulus* contra cocos Gram-positivos. **Scientia Plena**, [s.l.], v. 19, n. 8, p. 1-12, set. 2023.

FISCHBACH, Michael A.; WALSH, Christopher T.. Antibiotics for Emerging Pathogens. **Science**, [s.l.], v. 325, n. 5944, p. 1089-1093, Aug. 2009.

FRANCISCATO, Lidaiane Mariáh Silva dos Santos *et al.* Antimicrobial activity of cinnamon (*Cinnamomum verum*) essential oil and cinnamaldehyde against *Staphylococcus aureus*. **Research, Society And Development**, [s.l.], v. 11, n. 13, p. 1-17, Oct. 2022.

GALGANO, M. *et al.* Acquired bacterial resistance to antibiotics and resistance genes: from past to future. **Antibiotics**, [s.l.], v. 14, n. 3, p. 222, 2025.

GELATTI, Luciane Cristina *et al.* *Staphylococcus aureus* resistentes à metilicina: disseminação emergente na comunidade. **Anais Brasileiros de Dermatologia**, [s.l.], v. 84, n. 5, p. 501-506, out. 2009.

GONZÁLEZ, S. *et al.* Antimicrobial and antioxidant activities of essential oils from *Thymus vulgaris* and *Origanum vulgare* against foodborne pathogens. **Molecules**, [s.l.], v. 29, n. 2, art. 435, 2024.

ION, Adelina *et al.* Artificial neural network designed to identify NBOME hallucinogens based on 3D- MoRSE descriptors and topological descriptors. **2019 23rd International Conference On System Theory, Control And Computing (ICSTCC)**, [s.l.], p. 872-876, Oct. 2019.

IMELOUANE, B. *et al.* Chemical Composition and Antimicrobial Activity of Essential Oil of Thyme (*Thymus vulgaris*) from Eastern Morocco. **International Journal Of Agriculture And Biology**, [s.l.], v. 11, n. 2, p. 205-208, Apr. 2009.

JAMSHIDNEZHAD, Amir *et al.* The impact coenzyme Q10 supplementation on the inflammatory indices of women with breast cancer using A *machine learning* prediction model. **Informatics In Medicine Unlocked**, [s.l.], v. 24, p. 1-6, May 2021.

JOHN, Lijo *et al.* Chemoinformatics and Machine Learning Approaches for Identifying Antiviral Compounds. **Molecular Informatics**, [s.l.], v. 41, n. 4, p. 1-17, Nov. 2021.

KAUR, Gaganjot; CHHABRA, Amit. Improved J48 Classification Algorithm for the Prediction of Diabetes. **International Journal Of Computer Applications**, [s.l.], v. 98, n. 22, p. 13-17, July 2014.

KATCHBORIAN-NETO, Albert *et al.* Integrative open workflow for confident annotation and molecular networking of metabolomics MSE/DIA data. **Briefings In Bioinformatics**, [s.l.], v. 25, n. 2, p. 1-18, Jan. 2024.

KHAN, P.M.; ROY, K.. QSPR modelling for investigation of different properties of aminoglycoside-derived polymers using 2D descriptors. **SAR and QSAR in Environmental Research**, [s.l.], v. 32, n. 7, p. 595-614, June 2021.

KHOSRAVI, Mehran *et al.* Peppermint oil in irritable bowel syndrome: A systematic review and meta-analysis. **Journal of Clinical Gastroenterology**, [s.l.], v. 54, n. 9, p. 724–729, 2020.

KURNIAWAN, I.; ROSALINDA, M.; IKHSAN, N.. Implementation of ensemble methods on QSAR Study of NS3 inhibitor activity as anti-dengue agent. **SAR and QSAR in Environmental Research**, [s.l.], v. 31, n. 6, p. 477-492, June 2020.

- MARASCHIN, Marcelo *et al.* Metabolômica e quimiometria como ferramentas para análises quimio(bio) diversas. In: SOCCOL, Carlos Ricardo *et al.* **Biotecnologia Aplicada à Agro&Indústria**. São Paulo: Blucher, 2017. Cap. 1. p. 17-49. (5566).
- MENEZES, Renata de; SCOTTI, Luciana; SCOTTI, Marcus T.. Aprendizado de máquina aplicado a QSAR. **Química Nova**, [s.l.], v. 47, n. 7, p. 1-16, mar. 2024.
- MITRA, Susweta das *et al.* Essential oil and phytoconstituent (Linalool) from *Homalomena aromatica* Schott. rhizomes exhibit antibacterial and synergistic effects with beta-lactam antibiotics against Carbapenem-resistant *Enterobacteriaceae* (CRE) and Methicillin Resistant *S. aureus* (MRSA) pathogens. **Industrial Crops And Products**, [s.l.], v. 198, p. 116666, Aug. 2023.
- NEWMAN, D. J.; CRAGG, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. **Journal of Natural Products**, [s.l.], v. 83, n. 3, p. 770–803, 2020.
- ORDENES, Francisco Villarroel; SILIPO, Rosaria. *Machine learning* for marketing on the KNIME Hub: the development of a live repository for marketing applications. **Journal Of Business Research**, [s.l.], v. 137, p. 393-410, Dec. 2021.
- PENTEADO, Ana *et al.* Avaliação in vitro de Atividade Antimicrobiana de Óleos Essenciais contra *Salmonella typhimurium* e *Staphylococcus aureus*. **Higiene Alimentar**, [s.l.], v. 35, n. 293, p. 1-19, 2021.
- PEZANTES-ORELLANA, Camila *et al.* Essential oils: a systematic review on revolutionizing health, nutrition, and omics for optimal well-being. **Frontiers In Medicine**, [s.l.], v. 11, p. 1-17, Feb. 2024.
- PILON, Alan *et al.* Metabolômica de plantas: métodos e desafios. **Química Nova**, [s.l.], v. 43, n. 3, p. 329-354, mar. 2020.
- PINGAEW, Ratchanok *et al.* Anticancer activity and QSAR study of sulfur-containing thiourea and sulfonamide derivatives. **Heliyon**, [s.l.], v. 8, n. 8, p. 1-10, Aug. 2022.
- SAR, Taner *et al.* Novel Perspectives on Food-Based Natural Antimicrobials: a review of recent findings published since 2020. **Microorganisms**, [s.l.], v. 11, n. 9, p. 1-20, Sept. 2023.
- SKROVANKOVA, Sona; MLCEK, Jiri. Antioxidant Potential and Its Changes Caused by Various Factors in Lesser-Known Medicinal and Aromatic Plants. **Horticulturae**, [s.l.], v. 11, n. 1, p. 1-29, Jan. 2025.
- SWAMY, Mallappa Kumara; AKHTAR, Mohd Sayeed; SINNIHAH, Uma Rani. Antimicrobial Properties of Plant Essential Oils against Human Pathogens and Their Mode of Action: an updated review. **Evidence-Based Complementary And Alternative Medicine**, [s.l.], v. 2016, n. 1, p. 1-21, Jan. 2016.

VAIDYA, Ankur *et al.* Quantitative Structure-Activity Relationships: a novel approach of drug design and discovery. **Journal Of Pharmaceutical Sciences And Pharmacology**, [s.l.], v. 1, n. 3, p. 219-232, Sept. 2014.

VISWANATHAN, Anitha Kuttae; KRISHNAN, Rajkumar. Antimicrobial role of *Lavandula angustifolia* towards *Candida albicans*, *Streptococcus mutans*, *Staphylococcus aureus* and anti-adherence effect on denture base resin. **Journal Of Oral Biology And Craniofacial Research**, [s.l.], v. 14, n. 6, p. 815-824, Nov. 2024.

WANG, Mingxun *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. **Nature Biotechnology**, [s.l.], v. 34, n. 8, p. 828-837, Aug. 2016.

WORLD HEALTH ORGANIZATION. Antimicrobial resistance. Geneva: WHO, 2023. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>. Acesso em: jun. 2024.

ZAGO, Juliana A. A. *et al.* Sinergismo entre óleos essenciais e drogas antimicrobianas sobre linhagens de *Staphylococcus aureus* e *Escherichia coli* isoladas de casos clínicos humanos. **Revista Brasileira de Farmacognosia**, [s.l.], v. 19, n. 4, p. 828-833, dez. 2009.

ZHAN, Xin; YU, Shixiao. Reconstructing the historical patterns of forest stand based on CA-AdaBoost-ANN model. **Forest Ecology And Management**, [s.l.], v. 478, p. 1-9, Dec. 2020.

ZHANG, Shuxing *et al.* A Novel Automated Lazy Learning QSAR (ALL-QSAR) Approach: method development, applications, and virtual screening of chemical databases using validated all-qsar models. **Journal Of Chemical Information And Modeling**, [s.l.], v. 46, n. 5, p. 1984-1995, Sept. 2006.

ZHANG, Wen-Quan *et al.* Metabonomic analysis of the anti-inflammatory effects of volatile oils of *Angelica sinensis* on rat model of acute inflammation. **Biomedical Chromatography**, [s.l.], v. 29, n. 6, p. 902-910, Dec. 2014.

APÊNDICE A – Referências para coleta de dados do Modelo Lamiaceae vs. *S. aureus*

ALIMPIC, Ana *et al.* Composition and biological effects of *Salvia ringens* (Lamiaceae) essential oil and extracts. **Industrial Crops and Products**, [s.l.], v. 76, p. 702–709, 2015.

BEATOVIĆ, Damir *et al.* Chemical composition, antioxidant and antimicrobial activities of the essential oils of twelve *Ocimum basilicum* L. cultivars grown in Serbia. **Rec. Nat. Prod.**, [s.l.], v. 9, n. 1, p. 62–75, 2015.

BEJAOUI, Afef *et al.* Essential oil composition and antibacterial activity of *Origanum vulgare* subsp. *glandulosum* Desf. at different phenological stages. **Journal of Medicinal Food**, [s.l.], v. 16, n. 12, p. 1115–1120, 2013.

BENABED, Khadidja Houda *et al.* Chemical composition, antioxidant and antimicrobial activities of the essential oils of three Algerian Lamiaceae species. **Current Nutrition & Food Science**, [s.l.], v. 13, n. 2, p. 97–109, 2017.

CAO, Li *et al.* Essential oil composition, antimicrobial and antioxidant properties of *Mosla chinensis* Maxim. **Food Chemistry**, [s.l.], v. 115, p. 801–805, 2009.

CELIK TAS, O. Yesil *et al.* Antimicrobial activities of methanol extracts and essential oils of *Rosmarinus officinalis*, depending on location and seasonal variations. **Food Chemistry**, [s.l.], v. 100, p. 553–559, 2007.

DELAMARE, Ana Paula Longaray *et al.* Antibacterial activity of the essential oils of *Salvia officinalis* L. and *Salvia triloba* L. cultivated in South Brazil. **Food Chemistry**, [s.l.], v. 100, p. 603–608, 2007.

EBRAHIMABADI, Abdolrasoul H. *et al.* Essential oil composition and antioxidant and antimicrobial properties of the aerial parts of *Salvia eremophila* Boiss. from Iran. **Food and Chemical Toxicology**, [s.l.], v. 48, p. 1371–1376, 2010.

FATMA, Guesmi *et al.* In-vitro assessment of antioxidant and antimicrobial activities of methanol extracts and essential oil of *Thymus hirtus* sp. *algeriensis*. **Lipids in Health and Disease**, [s.l.], v. 13, p. 114, 2014.

GRUJIC-JOVANOVIC, Slavica *et al.* Composition and antibacterial activity of the essential oil of six *Stachys* species from Serbia. **Flavour and Fragrance Journal**, [s.l.], v. 19, p. 139–144, 2004.

HAJLAOUI, Hamed *et al.* Chemical composition and in vitro evaluation of antioxidant, antimicrobial, cytotoxicity and anti-acetylcholinesterase properties of Tunisian *Origanum majorana* L. essential oil. **Microbial Pathogenesis**, [s.l.], v. 95, p. 86–94, 2016.

HUSSAIN, Abdullah I. *et al.* Antibacterial activity of some Lamiaceae essential oils using resazurin as an indicator of cell growth. *LWT - Food Science and Technology*, [s.l.], v. 44, p. 1199–1206, 2011.

HUSSAIN, Abdullah I. *et al.* Seasonal variation in content, chemical composition and antimicrobial and cytotoxic activities of essential oils from four *Mentha* species. *Journal of the Science of Food and Agriculture*, [s.l.], v. 90, p. 1827–1836, 2010.

IMELUOANE, Amhamdi Hassan *et al.* Chemical composition and antimicrobial activity of essential oil of thyme (*Thymus vulgaris*) from Eastern Morocco. *International Journal of Agriculture and Biology*, [s.l.], v. 11, p. 205–208, 2009.

JIROVETZ, L. *et al.* Chemical composition, antimicrobial activities and odor descriptions of various *Salvia* sp. and *Thuja* sp. essential oils. *Ernährung/Nutrition*, [s.l.], v. 30, n. 4, p. 152-153, 2006.

KOUTSAVITI, Aikaterini *et al.* Antimicrobial activity and essential oil composition of five *Sideritis* taxa of *Empedoclia* and *Hesiodia* sect. from Greece. *Records of Natural Products*, [s.l.], v. 7, n. 1, p. 6–14, 2013.

LAWAL, Oladipupo A. *et al.* Chemical composition and antimicrobial activity of essential oil of *Ocimum kilimandscharicum* (R. Br.) Guerke: A new chemotype. *American Journal of Essential Oils and Natural Products*, [s.l.], v. 2, n. 1, p. 41–46, 2014.

LEMOS, Mayara Fumiere *et al.* Seasonal variation affects the composition and antibacterial and antioxidant activities of *Thymus vulgaris*. *Industrial Crops and Products*, [s.l.], v. 95, p. 543–548, 2017.

LIOLIOS, Christos *et al.* Chemical composition and antimicrobial activity of the essential oil of Algerian *Phlomis bovei* De Noé subsp. *bovei*. *Molecules*, [s.l.], v. 12, p. 772-781, 2007.

OKE, Feyza *et al.* Essential oil composition, antimicrobial and antioxidant activities of *Satureja cuneifolia* Ten. *Food Chemistry*, [s.l.], v. 112, p. 874–879, 2009.

OKOH, O.O. *et al.* Comparative evaluation of the antibacterial activities of the essential oils of *Rosmarinus officinalis* L. obtained by hydrodistillation and solvent-free microwave extraction methods. *Food Chemistry*, [s.l.], v. 120, p. 308–312, 2010.

PIEROZAN, Morgana Karen *et al.* Chemical characterization and antimicrobial activity of essential oils of *Salvia* L. species. *Ciência e Tecnologia de Alimentos*, [s.l.], v. 29, n. 4, p. 764-770, 2009.

RAPPER, Stephanie de *et al.* The in vitro antimicrobial effects of *Lavandula angustifolia* essential oil in combination with conventional antimicrobial agents. *Evidence-Based Complementary and Alternative Medicine*, [s.l.], v. 2016, Article ID 2752739, p. 1-9, 2016.

RUNYORO, D. *et al.* Chemical composition and antimicrobial activity of the essential oils of four *Ocimum* species growing in Tanzania. *Food Chemistry*, [s.l.], v. 119, p. 311–316, 2010.

SANDRI, I. G. *et al.* Antimicrobial activity of the essential oils of Brazilian species of the genus *Cunila* against foodborne pathogens and spoiling bacteria. **Food Chemistry**, [s.l.], v. 103, p. 823–828, 2007.

SARTORATTO, Adilson *et al.* Composition and antimicrobial activity of essential oils from aromatic plants used in Brazil. **Brazilian Journal of Microbiology**, Brazil, v. 35, p. 275–280, 2004.

SELLAM, Khalid *et al.* Biological investigations of antioxidant-antimicrobial properties and chemical composition of essential oil from *Lavandula multifida*. **Open Access Macedonian Journal of Medical Sciences**, [s.l.], v. 1, p. 1–7, 2013.

SHAKERI, Abolfazl *et al.* Chemical composition, antibacterial activity, and cytotoxicity of essential oil from *Nepeta ucrainica* L. spp. *kopetdaghensis*. **Industrial Crops and Products**, [s.l.], v. 58, p. 315–321, 2014.

SHARIFIFAR, F. *et al.* In vitro evaluation of antibacterial and antioxidant activities of the essential oil and methanol extract of endemic *Zataria multiflora* Boiss. **Food Control**, [s.l.], v. 18, p. 800–805, 2007.

STEFAN, Marius *et al.* The composition and antibacterial activity of essential oils in three *Ocimum* species growing in Romania. **Central European Journal of Biology**, [s.l.], v. 8, p. 600–608, 2013.

STOJKOVIĆ, Dejan *et al.* Chemical composition and antimicrobial activity of *Vitex agnus-castus* L. fruits and leaves essential oils. **Food Chemistry**, [s.l.], v. 128, p. 1017–1022, 2011.

TENORE, Gian Carlo *et al.* Antimicrobial and antioxidant properties of the essential oil of *Salvia lanigera* from Cyprus. **Food and Chemical Toxicology**, [s.l.], v. 49, p. 238–243, 2011.

TEPE, Bektas *et al.* Antimicrobial and antioxidant activities of the essential oil and various extracts of *Salvia tomentosa* Miller (*Lamiaceae*). **Food Chemistry**, [s.l.], v. 90, p. 333–340, 2005.

TOPLAN, Gizem Gulsoy *et al.* Composition and biological activities of *Salvia veneris* Hedge growing in Cyprus. **Industrial Crops and Products**, [s.l.], v. 97, p. 41–48, 2017.