

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE QUÍMICA

JONATHAN GONÇALVES MOTTA

FERRAMENTA INTERATIVA DE IDENTIFICAÇÃO DE COMPOSTOS
POR ANÁLISE PONDERADA DE ESPECTROS DE FT-IR: aplicação em agrotóxicos

RIO DE JANEIRO

2024

JONATHAN GONÇALVES MOTTA

**FERRAMENTA INTERATIVA DE IDENTIFICAÇÃO DE COMPOSTOS
POR ANÁLISE PONDERADA DE ESPECTROS DE FT-IR: aplicação em agrotóxicos**

Trabalho de conclusão de curso apresentado ao
Instituto de Química da Universidade Federal do Rio
de Janeiro, como parte dos requisitos necessários para
a obtenção do grau de Bacharel em Química.

Orientadores: Profa. Dra. Daniella Lopez Vale (DQA-IQ/UFRJ)
Prof. Dr. Ricardo Cunha Michel (DQA-IQ/UFRJ)

RIO DE JANEIRO

2024

CIP - Catalogação na Publicação

M921f Motta, Jonathan Gonçalves
Ferramenta interativa de identificação de compostos por análise ponderada de espectros de FT IR: aplicação em agrotóxicos / Jonathan Gonçalves Motta. -- Rio de Janeiro, 2024.
91 f.

Orientadora: Daniella Lopez Valle.
Coorientador: Ricardo Cunha Michel.
Trabalho de conclusão de curso (graduação) - Universidade Federal do Rio de Janeiro, Instituto de Química, Bacharel em Química, 2024.

1. Ferramenta de código aberto. 2. Algoritmo em Python. 3. Espectroscopia no infravermelho. 4. Agrotóxicos. I. Valle, Daniella Lopez, orient. II. Michel, Ricardo Cunha, coorient. III. Título.

JONATHAN GONÇALVES MOTTA

**FERRAMENTA INTERATIVA DE IDENTIFICAÇÃO DE COMPOSTOS
POR ANÁLISE PONDERADA DE ESPECTROS DE FT-IR: aplicação em agrotóxicos**

Trabalho de conclusão de curso apresentado ao
Instituto de Química da Universidade Federal do Rio
de Janeiro, como parte dos requisitos necessários para
a obtenção do grau de Bacharel em Química.

Profª. Dra. Daniella Lopez Vale, DQA-IQ/UFRJ

Prof. Dr. Ricardo Cunha Michel, DQA-IQ/UFRJ

Prof. Dr. Thiago Messias Cardozo, DFQ-IQ/UFRJ

Prof. Dr. Roberto Salgado Amado, DQI-IQ/UFRJ

Rio de Janeiro, RJ – Brasil

Agosto 2024

AGRADECIMENTOS

Aos meus pais e meus irmãos, que sempre me apoiaram a todo instante através dos dias longos, das noites cansativas e das madrugadas intermináveis de trabalho e estudo.

À Zuleide, minha avó materna, que me guiou e segurou minha mão ao longo de toda a minha infância, adolescência e vida adulta.

À Marisa, minha avó paterna, que, embora tenha partido cedo, deixou em mim a inspiração para ser uma pessoa melhor.

À Isabella, a pessoa mais bondosa, carinhosa e amiga que tive o prazer de ter como companheira, e com quem pude compartilhar tantos momentos e memórias. Mesmo que não esteja mais entre nós, sempre irei guardá-la em mim.

A todos os meus colegas e amigos próximos, em especial Raphael, Guilherme e Matheus, pelo apoio e auxílio indispensáveis na realização deste trabalho.

Ao Professor Roberto, meu coordenador de curso, por acreditar no meu potencial quando eu mesmo não acreditava.

Aos meus orientadores, Professor Michel e Professora Daniella, pelo suporte ao me permitir mudar de projeto para conseguir continuar no laboratório, e por serem sempre atenciosos e solícitos diante dos obstáculos que surgiram ao longo desse trabalho.

RESUMO

Este trabalho apresenta o desenvolvimento de uma ferramenta interativa para a identificação de compostos químicos por meio da análise ponderada de espectros de infravermelho por transformada de Fourier (FT-IR), com ênfase na análise de agrotóxicos. A espectroscopia no infravermelho é uma técnica analítica que permite a identificação de compostos químicos através da interação da radiação infravermelha com as moléculas dos compostos. Apesar de sua precisão e caráter não destrutivo, a interpretação dos espectros pode ser complexa e demorada. A ferramenta proposta visa facilitar a identificação de compostos de forma rápida e precisa, utilizando algoritmos de classificação desenvolvidos em Python, com o auxílio de espectros IV coletados de um banco de dados de acesso livre. A metodologia inclui a coleta, pré-processamento e armazenamento dos espectros em um banco de dados SQL, a implementação de algoritmos de análise e a construção de uma interface de usuário intuitiva desenvolvida em React.js e CSS. No final do processo, a ferramenta gera um relatório detalhado que inclui uma comparação gráfica e tabular dos cinco principais candidatos identificados como mais semelhantes ao composto analisado. O desempenho da ferramenta é comparado diretamente com a capacidade de classificação de algoritmos de aprendizado de máquina como K-Nearest Neighbors (KNN), Regressão Logística e Naive-Bayes. O objetivo principal é contribuir para o monitoramento ambiental, oferecendo uma solução acessível e eficiente para a identificação de agrotóxicos, facilitando o trabalho de pesquisadores e profissionais da área ambiental.

Palavras-chave: Agrotóxicos; Praguicidas; Espectroscopia no infravermelho; Classificação de espectros; Ferramenta de código aberto; Python.

ABSTRACT

This work presents the development of an interactive tool for the identification of chemical compounds through the weighted analysis of Fourier Transform Infrared (FT-IR) spectra, with an emphasis on the analysis of agrochemicals. Infrared spectroscopy is an analytical technique that allows the identification of chemical compounds through the interaction of infrared radiation with the molecules of the compounds. Despite its precision and non-destructive nature, the interpretation of spectra can be complex and time-consuming. The proposed tool aims to facilitate the identification of compounds quickly and accurately, using classification algorithms developed in Python and IR spectra collected from an open-access database. The methodology includes the collection, preprocessing, and storage of spectra in an SQL database, the implementation of analysis algorithms, and the construction of an intuitive user interface developed in React.js and CSS. At the end, the tool generates a detailed report that includes a graphical and tabular comparison of the top five candidates identified as most similar to the analyzed compound. The tool's performance is directly compared with the classification capability of machine learning algorithms such as K-Nearest Neighbors (KNN), Logistic Regression, and Naive Bayes. The main objective is to contribute to environmental monitoring by offering an accessible and efficient solution for the identification of agrochemicals, facilitating the work of researchers and environmental professionals.

Keywords: Agrochemicals; Pesticides; Infrared spectroscopy; Spectra classification; Open-source tools; Python

LISTA DE FIGURAS

Figura 1 – Espectro IV do composto de Ácido 1-naftalenoacético.	17
Figura 2 – Esquemático simplificado de um espectrômetro FT-IR.	18
Figura 3 – Grupos funcionais com diferentes características de absorção.	19
Figura 4 – Espectro IV do Ácido 1-naftalenoacético dividido em (a) 4000 – 2000 cm^{-1} e (b) 2000 – 400 cm^{-1} devidamente contornado.	30
Figura 5 – Fluxograma simplificado da metodologia API Rest.	38
Figura 6 – Interface interativa da tela inicial do usuário (Home).	39
Figura 7 – Interface interativa da tela de relatórios.	40
Figura 8 – (a) Informações resumidas sobre a análise do composto. (b) Espectros IV. (c) Tabela contendo as bandas de interesse. As páginas seguintes repetem o padrão de (b) e (c) para os outros componentes.	41
Figura 9 – Representação da estrutura química do Ácido 1-naftalenoacético.	42
Figura 10 – Espectros IV sobrepostos do Ácido 1-naftalenoacético do banco de dados (vermelho) e de entrada (preto).	43
Figura 11 – Classificação do Ácido 1-naftalenoacético por ordem de similaridade (maior para menor).	45
Figura 12 – Representação da estrutura química da Azociclotina.	45
Figura 13 – Espectros IV sobrepostos da Azociclotina do banco de dados (vermelho) e de entrada (preto).	46
Figura 14 – Classificação da Azociclotina por ordem de similaridade (maior para menor). ...	47
Figura 15 – Representação da estrutura química do Bendiocarbe.	48
Figura 16 – Espectros IV sobrepostos do Bendiocarbe do banco de dados (vermelho) e de entrada (preto).	49
Figura 17 – Classificação da Bendiocarbe por ordem de similaridade (maior para menor). ...	50
Figura 18 – Representação da estrutura química do Binapacril.	51
Figura 19 – Espectros IV sobrepostos do Binapacril do banco de dados (vermelho) e de entrada (preto).	52
Figura 20 – Classificação do Binapacril por ordem de similaridade (maior para menor).	53
Figura 21 – Representação da estrutura química do Coumafós.	54
Figura 22 – Espectros IV sobrepostos do Coumafós do banco de dados (vermelho) e de entrada (preto).	54
Figura 23 – Classificação do Coumafós por ordem de similaridade (maior para menor).	56

Figura 24 – Representação da estrutura química do Sulfato de endossulfam.	56
Figura 25 – Espectros IV sobrepostos do sulfato de endossulfam do banco de dados (vermelho) e de entrada (preto).	57
Figura 26 – Classificação do sulfato de endossulfam por ordem de similaridade (maior para menor).....	59
Figura 27 – Representação da estrutura química do Fention.....	59
Figura 28 – Espectros IV sobrepostos do Fention do banco de dados (vermelho) e de entrada (preto).	60
Figura 29 – Classificação do Fention por ordem de similaridade (maior para menor).	61
Figura 30 – Representação da estrutura química do Lenacil.....	62
Figura 31 – Espectros IV sobrepostos do Lenacil do banco de dados (vermelho) e de entrada (preto).	62
Figura 32 – Classificação do Lenacil por ordem de similaridade (maior para menor).	64
Figura 33 – Representação da estrutura química da Fosalone.....	64
Figura 34 – Espectros IV sobrepostos da Fosalone do banco de dados (vermelho) e de entrada (preto).	65
Figura 35 – Classificação da Fosalone por ordem de similaridade (maior para menor).	67
Figura 36 – Representação da estrutura química do Triadimefom.....	67
Figura 37 – Espectros IV sobrepostos do Triadimefom do banco de dados (vermelho) e de entrada (preto).....	68
Figura 38 – Classificação do Triadimefom por ordem de similaridade (maior para menor). ..	70

LISTA DE TABELAS

Tabela 1 – Resumo da estrutura dos dados do espectro IV do Ácido 1-naftalenoacético após uso do WebPlotDigitizer.....	31
Tabela 2 – Estrutura exemplo resumida utilizada para a análise com os algoritmos de aprendizado de máquina.....	33
Tabela 3 – Comparação das bandas de interesse entre o Ácido 1-naftalenoacético do banco de dados (Banco de dados) e o de entrada (Amostra).	44
Tabela 4 – Comparação das bandas de interesse entre a Azociclotina do banco de dados (Banco de dados) e o de entrada (Amostra).	46
Tabela 5 – Comparação das bandas de interesse entre o Bendiocarbe do banco de dados (Banco de dados) e o de entrada (Amostra).	49
Tabela 6 – Comparação das bandas de interesse entre o Binapacril do banco de dados (Banco de dados) e o de entrada (Amostra).	52
Tabela 7 – Comparação das bandas de interesse entre o Coumafós do banco de dados (Banco de dados) e o de entrada (Amostra).	55
Tabela 8 – Comparação das bandas de interesse entre o sulfato de endossulfam do banco de dados (Banco de dados) e o de entrada (Amostra).	57
Tabela 9 – Comparação das bandas de interesse entre o Fention do banco de dados (Banco de dados) e o de entrada (Amostra).	60
Tabela 10 – Comparação das bandas de interesse entre o Lenacil do banco de dados (Banco de dados) e o de entrada (Amostra).	63
Tabela 11 – Comparação das bandas de interesse entre a Fosalone do banco de dados (Banco de dados) e o de entrada (Amostra).	65
Tabela 12 – Comparação das bandas de interesse entre o Triadimefom do banco de dados (Banco de dados) e o de entrada (Amostra).	68

LISTA DE QUADROS

Quadro 1 – Classificação do potencial de periculosidade ambiental.	21
Quadro 2 – Classificação toxicológica obsoleta de 1992.	21
Quadro 3 – Reclassificação toxicológica atualizada de 2019.	22

SUMÁRIO

1	INTRODUÇÃO	13
2	OBJETIVOS	14
2.1	OBJETIVO GERAL.....	14
2.2	OBJETIVOS ESPECÍFICOS	14
3	FUNDAMENTAÇÃO TEÓRICA.....	15
3.1	ESPECTROSCOPIA NO INFRAVERMELHO	15
3.1.1	Radiação Infravermelha	15
3.1.2	Lei de Lambert-Beer-Bouguer	16
3.1.3	Espectroscopia no infravermelho por transformada de Fourier	17
3.1.4	Interpretação de um espectro	18
3.2	AGROTÓXICOS	19
3.2.1	Definição e regulamentação	19
3.2.2	Classificação toxicológica.....	20
3.2.3	Impactos na saúde humana e no meio ambiente	22
3.2.3.1	<i>Consequências na saúde humana.....</i>	<i>23</i>
3.2.3.2	<i>Consequências no meio ambiente</i>	<i>24</i>
3.2.4	Aplicação da técnica de espectroscopia IV no monitoramento ambiental	24
3.3	APRENDIZADO DE MÁQUINA	25
3.3.1	K-Nearest Neighbors (KNN)	25
3.3.2	Regressão Logística	26
3.3.3	Naive-Bayes	26
3.3.4	Aplicações de aprendizado de máquina em análises no infravermelho	27
4	METODOLOGIA.....	28
4.1	COLETA, PRÉ-TRATAMENTO DOS ESPECTROS E EXTRAÇÃO DOS DADOS NUMÉRICOS.....	28
4.1.1	Descrição da fonte dos dados.....	28
4.1.2	Procedimentos de coleta.....	29
4.1.3	Pré-processamento dos espectros IV.....	29
4.1.4	Conversão do espectro IV de imagem para dados numéricos.....	30
4.1.4.1	<i>Extração dos dados quantitativos.....</i>	<i>30</i>
4.1.4.2	<i>Correção da linha de base</i>	<i>31</i>
4.2	DESENVOLVIMENTO DO BANCO DE DADOS SQL	32

4.2.1	Configuração para o algoritmo principal.....	32
4.2.2	Configuração para os algoritmos de aprendizado de máquina.....	32
4.3	ARQUITETURA E FUNCIONAMENTO DO ALGORITMO.....	33
4.3.1	Filtragem das bandas de interesse	33
4.3.2	Algoritmo de similaridade	34
4.4	BIBLIOTECAS DE ALGORITMOS DE APRENDIZADO DE MÁQUINA	35
4.4.1	Normalização dos dados.....	35
4.4.2	Modelos de classificação por aprendizado de máquina	36
5	RESULTADOS E DISCUSSÕES	37
5.1	ESTRUTURA DO BANCO DE DADOS.....	37
5.2	CRITÉRIOS DE SELEÇÃO DAS AMOSTRAS PARA ANÁLISE.....	37
5.3	INTERFACE DE USUÁRIO E INTEGRAÇÃO DA ARQUITETURA	38
5.4	CONCEPÇÃO E DESIGN DO RELATÓRIO DE SAÍDA	40
5.5	ANÁLISE DAS AMOSTRAS	42
5.5.1	Ácido 1-naftalenoacético	42
5.5.2	Azociclotina	45
5.5.3	Bendiocarbe.....	48
5.5.4	Binapacril	51
5.5.5	Coumafós.....	54
5.5.6	Sulfato de endossulfam.....	56
5.5.7	Fention	59
5.5.8	Lenacil	62
5.5.9	Fosalona.....	64
5.5.10	Triadimefom	67
6	CONCLUSÕES E CONSIDERAÇÕES FINAIS	71
7	SUGESTÕES DE TRABALHOS FUTUROS.....	72
8	REFERÊNCIAS	73
	APÊNDICE A – COMPONENTES DO BANCO DE DADOS	78
	APÊNDICE B – ESTRUTURAS DOS RESULTADOS DO ÁCIDO 1-NAFTALENOACÉTICO	80
	APÊNDICE C – ESTRUTURAS DOS RESULTADOS DA AZOCICLOTINA.....	81
	APÊNDICE D – ESTRUTURAS DOS RESULTADOS DO BENDIOCARBE	82
	APÊNDICE E – ESTRUTURAS DOS RESULTADOS DO BINAPACRIL	83
	APÊNDICE F – ESTRUTURAS DOS RESULTADOS DO COUMAFÓS	84

APÊNDICE G – ESTRUTURAS DOS RESULTADOS DO SULFATO DE ENDOSSULFAM	85
APÊNDICE H – ESTRUTURAS DOS RESULTADOS DO FENTION.....	86
APÊNDICE I – ESTRUTURAS DOS RESULTADOS DO LENACIL	87
APÊNDICE J – ESTRUTURAS DOS RESULTADOS DA FOSALONA	88
APÊNDICE K – ESTRUTURAS DOS RESULTADOS DO TRIADIMEFOM.....	89

1 INTRODUÇÃO

Agrotóxicos são substâncias químicas usadas na agricultura para controlar organismos indesejáveis que podem comprometer a produção agrícola. Este termo engloba uma variedade de produtos, incluindo herbicidas, fungicidas, pesticidas, rodenticidas, entre outros. No Brasil, a definição e regulamentação dos agrotóxicos eram regidas pela Lei Federal nº 7.802/1989 em conjunto com o Decreto nº 4.074/2002, até a entrada da Lei Federal nº 14.785/2023 atualmente em vigor. Esses produtos são essenciais para a proteção das culturas agrícolas, no entanto, apresentam riscos significativos para a saúde humana e o meio ambiente (Santana; Moura; Nogueira, 2013).

A exposição extensiva a agrotóxicos no Brasil e no mundo gera uma série de preocupações. Segundo Do Nascimento *et al.* (2020), os principais problemas incluem impactos negativos na saúde humana, como intoxicações agudas e crônicas que resultam em uma variedade de sintomas. Além disso, há a contaminação de solos e recursos hídricos, levando a uma possível degradação ambiental irreversível. Dados epidemiológicos estimam que de três a cinco milhões de pessoas são contaminadas por agrotóxicos a cada ano em todo o mundo, destacando a necessidade urgente de monitoramento e controle rigoroso dessas substâncias.

Uma técnica poderosa utilizada para a identificação e monitoramento de agrotóxicos é a espectroscopia no infravermelho (IV). A espectroscopia IV é uma técnica analítica que permite a identificação de compostos químicos através da análise das interações entre a radiação infravermelha e as moléculas dos compostos. Essa técnica é fundamental para o estudo e monitoramento de agrotóxicos, pois oferece uma forma precisa, não destrutiva e ecológica de analisar amostras ambientais. No entanto, a interpretação dos espectros pode ser complexa e demorada, exigindo ferramentas adequadas para facilitar o processo (Carvalho *et al.*, 2023).

A análise manual dos espectros pode ser suscetível a erros, afetando a precisão dos resultados. Isso limita a eficiência e a acessibilidade da técnica para pesquisas e aplicações práticas. Além disso, a falta de ferramentas interativas e acessíveis que integrem bancos de dados de espectros e algoritmos de análise representa uma barreira significativa. Atualmente, as opções disponíveis são geralmente complexas, caras e inacessíveis para muitos usuários, especialmente em contextos acadêmicos e de pesquisa ambiental.

Com o objetivo de superar essas limitações, este trabalho propõe o desenvolvimento de uma ferramenta interativa para a identificação de compostos por meio da análise ponderada de espectros de FT-IR. Esta iniciativa busca contribuir para o monitoramento ambiental, fornecendo uma solução acessível e eficaz para a identificação de agrotóxicos.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Desenvolver uma ferramenta interativa que auxilie na identificação de compostos químicos através da análise ponderada de espectros de FT-IR (Espectroscopia de Infravermelho por Transformada de Fourier). A ferramenta deverá ser de fácil utilização, integrando algoritmos de classificação desenvolvidos em Python com um banco de dados de acesso livre, para proporcionar resultados rápidos e precisos. Ao final do processo, o usuário receberá um relatório detalhado que inclui comparações gráficas e tabulares dos compostos identificados, visando melhorar a eficiência, a precisão e a acessibilidade das análises espectroscópicas em contextos acadêmicos e de pesquisa ambiental.

2.2 OBJETIVOS ESPECÍFICOS

- Extrair as informações gráficas dos espectros no infravermelho de agrotóxicos provenientes do banco de dados de livre acesso, processar esses dados e armazenar em um banco de dados em linguagem SQL, para facilitar o acesso e a manipulação dos dados;
- Implementar o método de pré-processamento nos dados de entrada fornecidos pelo usuário para garantir a compatibilidade e precisão na análise dos espectros;
- Desenvolver a lógica do algoritmo de classificação e construir uma interface de usuário intuitiva e amigável, utilizando React.js, CSS, e Python;
- Comparar o desempenho do algoritmo desenvolvido com algoritmos de aprendizado de máquina comumente utilizados para classificação de dados, avaliando sua eficácia e precisão na identificação de compostos químicos;
- Fornecer um relatório detalhado ao usuário, incluindo comparações gráficas e tabulares dos cinco principais candidatos identificados, facilitando a interpretação dos resultados e a tomada de decisões.

3 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica deste trabalho aborda os principais conceitos e técnicas envolvidos na identificação de compostos químicos por espectroscopia no infravermelho (IV) e sua análise por algoritmos de aprendizado de máquina. Primeiramente, são discutidos os princípios fundamentais da espectroscopia IV, incluindo a natureza da radiação infravermelha, a Lei de Lambert-Beer-Bouguer e a espectroscopia por transformada de Fourier (FT-IR). Em seguida, é apresentada uma visão geral sobre os agrotóxicos, abordando sua definição, regulamentação e os impactos na saúde humana e no meio ambiente. A aplicação da espectroscopia IV no monitoramento ambiental é explorada, destacando-se suas vantagens e limitações. Por fim, são introduzidos os conceitos de aprendizado de máquina e os algoritmos específicos utilizados neste estudo, como K-Nearest Neighbors (KNN), Regressão Logística e Naive-Bayes, que são fundamentais para a classificação e análise dos espectros obtidos.

3.1 ESPECTROSCOPIA NO INFRAVERMELHO

A espectroscopia no infravermelho (IV) é uma técnica analítica essencial para a identificação e caracterização de compostos químicos. Baseada na interação da radiação infravermelha com as moléculas, essa técnica permite a análise das vibrações moleculares, fornecendo informações detalhadas sobre a estrutura e a composição dos materiais analisados. Este subitem explora os princípios básicos da espectroscopia IV, incluindo a radiação infravermelha, a Lei de Lambert-Beer-Bouguer, e a espectroscopia por transformada de Fourier (FT-IR), destacando sua importância e aplicação na química analítica.

3.1.1 Radiação Infravermelha

A radiação infravermelha (IV) encontra-se localizada entre a luz visível e as micro-ondas no espectro eletromagnético. É especialmente útil para a identificação de compostos na faixa entre 4000 e 400 cm^{-1} , sendo frequentemente empregada em estudos práticos. Além dessa região, existem outras áreas de interesse, como o infravermelho próximo (4000 até 14.290 cm^{-1}) e o infravermelho distante (700 até 200 cm^{-1}) (Silverstein *et al.*, 2015). Embora técnicas como a Ressonância Magnética Nuclear (RMN) e a espectrometria de massas tenham ganhado popularidade, a espectroscopia IV continua sendo uma ferramenta poderosa para a análise da estrutura molecular e composição dentro dessas faixas especificadas.

A região do espectro eletromagnético abaixo de 100 cm^{-1} , conhecida por sua associação com rotações moleculares puras, é predominantemente absorvida nas frequências de micro-ondas e IV distante. Em contraste, as vibrações moleculares, responsáveis por bandas de absorção na maior parte da região de IV, estendem-se desde aproximadamente 10.000 até 100 cm^{-1} (Colthup; Daly; Wiberley, 2012). Nesta faixa de absorção, espectros vibracionais aparecem como bandas ao invés de linhas devido ao fato que uma mudança de energia vibracional é diretamente acompanhada por um número de mudanças de energias rotacionais. Notavelmente, entre 4000 e 400 cm^{-1} , ocorrem as bandas vibracionais-rotacionais de interesse para análise dos compostos. A frequência de absorção, medida em número de onda na faixa de interesse, depende da massa dos átomos, da força das suas ligações e da sua configuração geométrica (Silverstein *et al.*, 2015).

3.1.2 Lei de Lambert-Beer-Bouguer

A intensidade das bandas vibracionais-rotacionais pode ser expressa em transmitância (T) ou absorvância (A). A lei de Lambert-Beer-Bouguer estabelece uma relação linear entre a absorvância de uma amostra e a concentração do analito. A equação que descreve essa relação é mostrada abaixo (Colthup; Daly; Wiberley, 2012):

$$A = \varepsilon bC \quad (1)$$

- A representa a absorvância,
- ε é denominado como coeficiente de absorvidade molar,
- b é o comprimento do caminho óptico do feixe de luz através da amostra,
- C é a concentração do analito.

Enquanto a transmitância é expressa como (Colthup; Daly; Wiberley, 2012):

$$T = \frac{I}{I_0} = 10^{-\varepsilon bC} \quad (2)$$

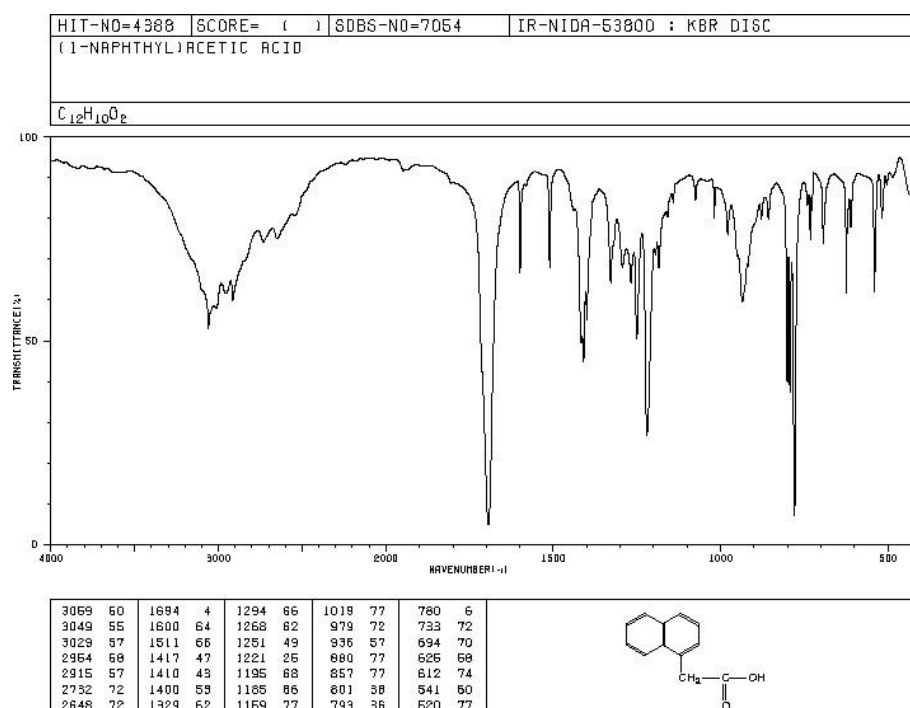
- T representa a transmitância,
- I é a intensidade da luz transmitida pela amostra,
- I_0 é a intensidade da radiação monocromática incidente na amostra.

Dessa forma, relaciona-se absorvância e transmitância da seguinte maneira:

$$A = \log_{10} \frac{1}{T} \text{ ou } A = \log_{10} \frac{100}{\%T} \quad (3)$$

A Figura 1 mostra um exemplo de um espectro IV expresso em porcentagem de número de onda (cm^{-1}) versus transmitância (%T).

Figura 1 – Espectro IV do composto de Ácido 1-naftalenoacético.



Fonte: Spectral Database for Organic Compounds (SDBS). **Espectro de Infravermelho; SDBS N°: 7054; IR-NIDA-53800.** Disponível em: <https://sdb.sdb.aist.go.jp/>. Acesso em: 30 jun. 2024.

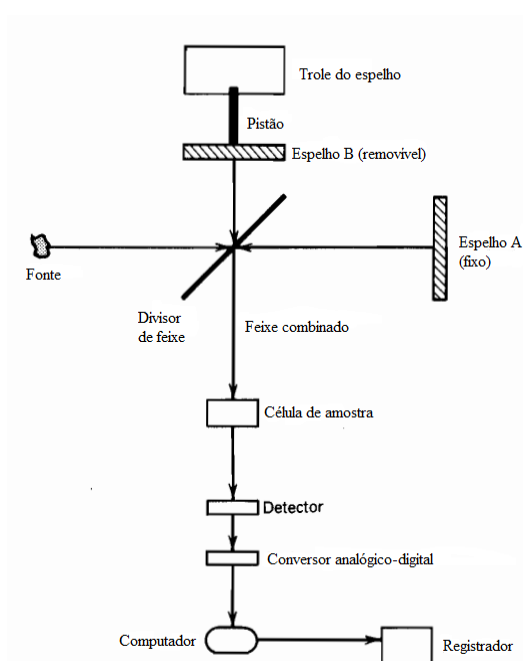
3.1.3 Espectroscopia no infravermelho por transformada de Fourier

Atualmente, a técnica de espectroscopia no infravermelho por transformada de Fourier (FT-IR) representa uma evolução significativa em relação às técnicas anteriores de espectroscopia IV. Tradicionalmente, os espectros eram obtidos por meio da passagem de um feixe de IV pela amostra, com o espectro sendo escaneado utilizando um dispositivo de dispersão, como grades de difração (Silverstein *et al.*, 2015).

Diferentemente das técnicas convencionais que dependem de um monocromador para selecionar as frequências de radiação uma a uma, a FT-IR permite que toda a faixa de radiação passe pela amostra simultaneamente. A resolução obtida por instrumentos FT-IR pode alcançar

até aproximadamente $0,001\text{ cm}^{-1}$. Além disso, a conversão analógico-digital facilita a integração dos resultados com diversas outras técnicas, como a cromatografia líquida de alta eficiência (HPLC) e a cromatografia gasosa (GC) (Silverstein *et al.*, 2015). A Figura 2 apresenta um esquemático simplificado de um espectrômetro FT-IR.

Figura 2 – Esquemático simplificado de um espectrômetro FT-IR.

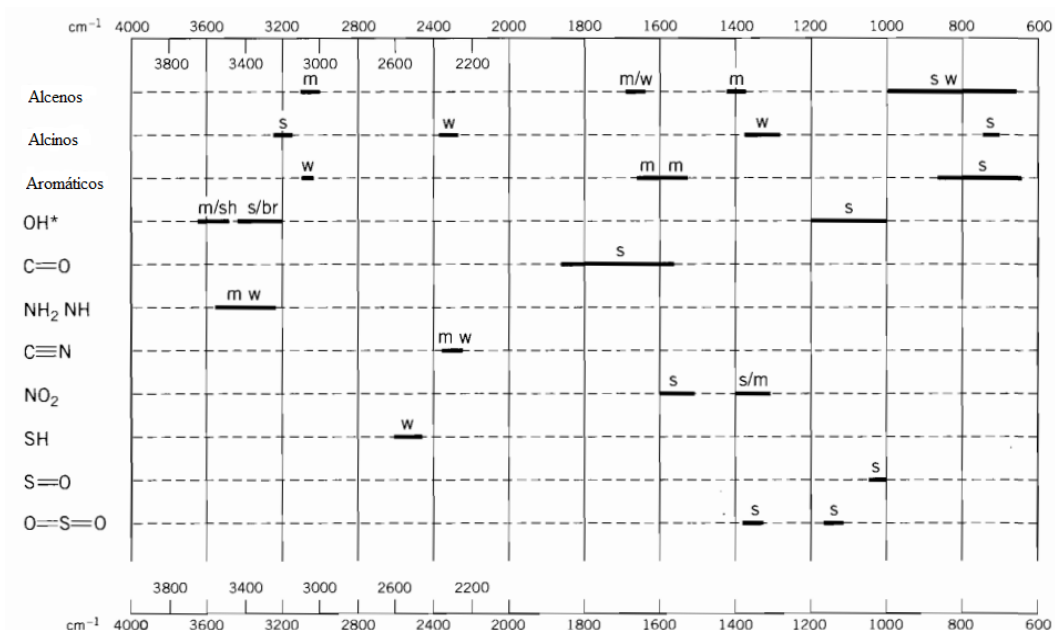


Fonte: Silverstein *et al.*, 2015. Adaptado pelo autor.

3.1.4 Interpretação de um espectro

Para uma interpretação eficaz de um espectro IV, deve-se seguir algumas diretrizes essenciais. Primeiramente, é necessário assegurar que o espectro esteja com a intensidade adequada e que a amostra analisada seja relativamente pura. Além disso, é fundamental que o espectrofotômetro esteja calibrado adequadamente e que o método utilizado para o manuseio da amostra seja claramente especificado. Nos espectros IV, os grupos funcionais das moléculas exibem faixas de absorção consistentes, o que permite compará-los através de diferentes espectros. As bandas de um espectro podem normalmente ser descritas por termos semiquantitativos (s = forte, m = média, w = fraca) (Silverstein *et al.*, 2015). Um exemplo ilustrativo dos padrões típicos de absorção de alguns grupos funcionais em termos semiquantitativos pode ser observado na Figura 3.

Figura 3 – Grupos funcionais com diferentes características de absorção.



Fonte: Silverstein *et al.*, 2015. Adaptado pelo autor.

3.2 AGROTÓXICOS

Os agrotóxicos são substâncias químicas empregadas na agricultura com o objetivo de controlar pragas e doenças que afetam as plantações, sendo essenciais para assegurar a produtividade agrícola e a qualidade dos alimentos. No entanto, seu uso implica em riscos significativos para a saúde humana e o meio ambiente, exigindo regulamentação rigorosa e práticas seguras de aplicação. Esta seção aborda as definições e classificações toxicológicas dos agrotóxicos, examina os impactos na saúde humana e no meio ambiente, e explora o uso da espectroscopia IV no monitoramento ambiental.

3.2.1 Definição e regulamentação

O uso do termo "agrotóxico" foi adotado oficialmente no Brasil através da Lei Federal nº 7.802/1989, posteriormente regulamentada pelo Decreto nº 4.074/2002, que estabeleceu as bases legais para a regulação desses produtos no país (Santana; Moura; Nogueira, 2013). Essa legislação define agrotóxicos e afins como:

- a) os produtos e os agentes de processos físicos, químicos ou biológicos, destinados ao uso nos setores de produção, no armazenamento e beneficiamento de produtos agrícolas, nas pastagens, na proteção de

florestas, nativas ou implantadas, e de outros ecossistemas e também de ambientes urbanos, hídricos e industriais, cuja finalidade seja alterar a composição da flora ou da fauna, a fim de preservá-las da ação danosa de seres vivos considerados nocivos (Brasil, 1989);

b) substâncias e produtos, empregados como desfolhantes, dessecantes, estimuladores e inibidores de crescimento (Brasil, 1989).

Conforme da Rocha, dos Santos e Pinto (2022) apontam, embora o termo tenha sido definido na legislação, este é frequentemente referido por diversas designações genéricas, não apenas variando semanticamente, sendo incluídas expressões como pesticidas, praguicidas, defensivos agrícolas, defensivos fitossanitários e fitossanitários.

Desde a adoção inicial da legislação, diversas atualizações foram implementadas, especialmente após a criação da Agência Nacional de Vigilância Sanitária (ANVISA), do Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA), e do Ministério da Agricultura, Pecuária e Abastecimento (MAPA). Esses órgãos sendo diretamente responsáveis pela regulamentação e fiscalização dos agrotóxicos no Brasil, conforme discutido por De Itoz e Pereira em 2024.

Em 28 de dezembro de 2023, mais de trinta anos após a primeira lei dos agrotóxicos, foi publicada e entrou em vigor a Lei Federal nº 14785/2023, que amplia a regulamentação, a produção, comercialização e uso de agrotóxicos no Brasil, revogando a Lei nº 7.802/1989 (De Itoz; Pereira, 2024).

O artigo 2º dessa nova legislação contém classificações significativamente importantes, pois apresenta cinquenta e dois incisos, evidenciando uma expansão considerável na regulamentação dos agrotóxicos e de produtos de controle ambiental (Brasil, 2023).

3.2.2 Classificação toxicológica

Como mencionado anteriormente, um marco regulatório estabelecido em 1989 instituiu a avaliação compartilhada dos registros de agrotóxicos por diferentes órgãos: o MAPA, encarregado de avaliar a eficácia agrônômica; o IBAMA, responsável pela análise dos impactos ambientais; e a ANVISA, que examina os riscos relacionados à toxicidade humana (da Rocha; dos Santos; Pinto, 2022).

A análise dos impactos ambientais e de toxicidade humana são conduzidas utilizando-se classificações organizadas em classes ou categorias. Em relação ao potencial de periculosidade ambiental, o Quadro 1 mostra a classificação baseada nos parâmetros de

bioacumulação, persistência, transporte, toxicidade a diversos organismos, potencial mutagênico, teratogênico e carcinogênico estabelecido pelo IBAMA (1996).

Quadro 1 – Classificação do potencial de periculosidade ambiental.

Classe	Grau
Classe I	Produto Altamente Perigoso
Classe II	Produto Muito Perigoso
Classe III	Produto Perigoso
Classe IV	Produto Pouco Perigoso

Fonte: Adaptado de IBAMA (1996).

Quanto à toxicidade humana, a Secretaria Nacional de Vigilância Sanitária (1992), definiu uma classificação em classes para os agrotóxicos. Nesta classificação, foram considerados estudos da chamada DL 50 (dose letal), que é a dose considerada necessária de uma dada substância para matar 50% de uma população em teste. Para que a determinação seja feita, expõe-se as cobaias a várias doses da substância até que seja encontrado o valor que mata apenas metade da população testada.

Os produtos comprovadamente teratogênicos, carcinogênicos ou mutagênicos eram excluídos de receber uma classificação toxicológica. Da mesma forma, produtos que causassem efeitos como corrosão, ulceração ou opacidade irreversível na córnea eram submetidos a estudos especiais para determinar se receberiam ou não uma classificação toxicológica (Secretaria Nacional de Vigilância Sanitária, 1992). O Quadro 2 ilustra essa classificação em melhores detalhes.

Quadro 2 – Classificação toxicológica obsoleta de 1992.

Classe	Grau	DL 50*
Classe I	Produtos Extremamente Tóxicos	$< 5 \text{ mg kg}^{-1}$
Classe II	Produtos Altamente Tóxicos	5 mg kg^{-1} a 50 mg kg^{-1}
Classe III	Produtos Medianamente Tóxicos	50 mg kg^{-1} a 500 mg kg^{-1}
Classe IV	Produtos Pouco Tóxicos	$> 500 \text{ mg kg}^{-1}$

* Formulações sólidas que apresentam DL 50 oral.

Fonte: Adaptado de Secretaria Nacional de Vigilância Sanitária (1992).

Em 2019, entrou em vigor a Resolução-RE nº 2080, conhecida como o Novo Marco Regulatório. Esta resolução expandiu o sistema anterior de quatro classes, para cinco categorias de classificação toxicológica, sendo incluído o item “não classificado” para produtos de baixíssimo potencial de dano (ANVISA, 2019). A nova classificação pode ser observada no Quadro 3, apresentado abaixo:

Quadro 3 – Reclassificação toxicológica atualizada de 2019.

Categoria	Grau
Categoria 1	Produto Extremamente Tóxico
Categoria 2	Produto Altamente Tóxico
Categoria 3	Produto Moderadamente Tóxico
Categoria 4	Produto Pouco Tóxico
Categoria 5	Produto Improvável de Causar Dano Agudo
Não classificado	Produto Não Classificado

Fonte: Adaptado de ANVISA (2019).

A justificativa apresentada para a reclassificação de toxicidade humana de agrotóxicos no Brasil foi de alinhar as práticas nacionais com as normativas adotadas em vários países da Ásia e da União Europeia. Essa proposta intencionava adotar as diretrizes de rotulagem do Sistema Globalmente Harmonizado de Classificação e Rotulagem de Produtos Químicos (*Globally Harmonized System of Classification and Labelling of Chemicals* – GHS). No entanto, mesmo com a implementação do sistema GHS, verifica-se que o Brasil não aderiu completamente a tais regulamentações, visto que mais de 30% dos pesticidas em circulação no país são proibidos na União Europeia, conforme apontado por Sampaio (2019).

3.2.3 Impactos na saúde humana e no meio ambiente

O uso extensivo de agrotóxicos tem sido associado a significativos danos cumulativos tanto para a saúde humana quanto para o meio ambiente. A Organização Mundial da Saúde (OMS) estima que, anualmente, de três a cinco milhões de pessoas globalmente são afetadas por contaminações relacionadas a agrotóxicos. Este número é ainda mais alarmante em países em desenvolvimento, onde se estima que até 25 milhões de trabalhadores rurais estão expostos a cada ano (Do Nascimento *et al.*, 2020; Peres; Moreira; Dubois, 2003).

A exposição imprópria e indiscriminada aos agrotóxicos representa um dos desafios mais sérios para a saúde pública. Os impactos nocivos dessas substâncias no meio ambiente e na saúde humana dependem de vários fatores, incluindo o princípio ativo do composto químico, a dose absorvida, a via de exposição, e as características individuais do exposto. Neste contexto, as alegações de uso seguro feitas pelas indústrias podem ocultar os verdadeiros riscos que esses produtos representam para os seres humanos (Do Nascimento *et al.*, 2020).

No Brasil, os critérios de avaliação da toxicidade e o risco dos agrotóxicos são estabelecidos conforme a Seção 3.2.2. No entanto, de acordo com a Organização Mundial da Saúde (2010) e outras diretrizes internacionais, os pesticidas considerados altamente perigosos (HHPs) são classificados segundo outros critérios específicos. Estes incluem a toxicidade aguda, posicionando os pesticidas nas classes de maior perigo da OMS, Ia ou Ib. E a toxicidade crônica, onde são consideradas a carcinogenicidade, a mutagenicidade e a toxicidade reprodutiva, que são categorizadas pela GHS nas Categorias 1A e 1B.

3.2.3.1 Consequências na saúde humana

O diagnóstico de contaminação por agrotóxicos enfrenta desafios significativos devido à variedade de substâncias químicas envolvidas e à complexidade técnica dos testes necessários, tornando-os pouco acessíveis em grande escala no Sistema Único de Saúde (SUS). Embora existam métodos capazes de identificar a presença de venenos no organismo, como os que detectam organofosforados e carbamatos, estes se limitam a uma fração dos agentes tóxicos disponíveis e muitas vezes só são eficazes em casos de intoxicação aguda. Além disso, o corpo humano pode não reter vestígios detectáveis desses venenos por longo prazo, complicando ainda mais a confirmação do diagnóstico (Londres, 2012).

Segundo Do Nascimento *et al.* (2020), as intoxicações agudas causadas por agrotóxicos podem manifestar-se através de sintomas como cefaleia, fraqueza, mal-estar geral, dor estomacal e sonolência. Além desses efeitos imediatos, a exposição prolongada a esses produtos químicos está associada a doenças crônicas, incluindo câncer, doença de Alzheimer, esclerose múltipla e doença de Parkinson. Adicionalmente, há evidências significativas que sugerem uma conexão direta entre a exposição a agrotóxicos e um aumento nas taxas de suicídio entre trabalhadores rurais, indicando que essas substâncias podem ter efeitos adversos graves sobre o sistema nervoso central, resultando em transtornos como ansiedade, irritabilidade, problemas de sono e depressão.

3.2.3.2 Consequências no meio ambiente

O uso principal dos agrotóxicos é aumentar a produtividade agrícola por meio de sua aplicação em culturas. No entanto, segundo Veiga *et al.* (2006), esses produtos químicos podem contaminar o solo e os sistemas hídricos, resultando em degradação ambiental significativa. Essa contaminação pode levar a danos duradouros nos ecossistemas, com efeitos potencialmente irreversíveis ao longo do tempo.

Os agrotóxicos são aplicados diretamente nas plantas ou no solo. No entanto, mesmo os que são destinados às plantas frequentemente acabam no solo, arrastados pela chuva ou pela irrigação. Essa transferência, especialmente em áreas com aquíferos naturalmente vulneráveis, pode tornar a água imprópria para consumo. Isso representa um risco significativo não apenas para a população local, mas também para todos aqueles que dependem dessa fonte de água, conforme destacado por Veiga *et al.* (2006). Além disso, Bohner, Araújo e Nishijima (2013) apontam que, dos agrotóxicos aplicados por pulverização, menos de 10% atingem efetivamente o alvo pretendido, aumentando ainda mais o risco de contaminação ambiental.

3.2.4 Aplicação da técnica de espectroscopia IV no monitoramento ambiental

A espectroscopia IV, com sua capacidade de fornecer resultados analíticos rápidos, possui diversas aplicações práticas. Na indústria de polímeros, por exemplo, ela é essencial para o controle de qualidade ao longo das linhas de produção, assegurando a consistência do material produzido. Além disso, a espectroscopia IV desempenha um papel crucial na indústria alimentícia, onde é utilizada para investigar possíveis adulterações na composição química dos alimentos, garantindo a segurança e a conformidade dos produtos (Bayu; Nandiyanto; Ragadhita, 2023; Mendes; Duarte, 2021).

Além de suas aplicações industriais, a espectroscopia IV também se mostra eficaz na área de monitoramento de pesticidas, como demonstrado por Grigoriu *et al.* (2018). Neste estudo, a técnica FTIR/ATR foi utilizada para monitorar a degradação do pesticida Dinocape em amostras de solo, vinho e uvas. Essa aplicação fica ainda mais evidente no trabalho de Carvalho *et al.* (2023), onde foi aplicada a técnica para estudar os efeitos de herbicidas à base de Diuron em feijões carioca, utilizando análises estatísticas multivariadas para detectar alterações nas folhas tratadas.

3.3 APRENDIZADO DE MÁQUINA

O aprendizado de máquina é uma área da inteligência artificial que permite que sistemas aprendam e melhorem a partir de experiências sem serem explicitamente programados. Esta seção discute principalmente três metodologias amplamente utilizadas e que serão aplicadas ao longo deste estudo: K-Nearest Neighbors (KNN), Regressão Logística e Naive-Bayes. Essas técnicas serão empregadas para resolver problemas de classificação e predição, que são cruciais para a análise de espectros IV no contexto deste trabalho.

As três técnicas mencionadas são exemplos de algoritmos de aprendizado supervisionado, ou seja, dependem de tarefas específicas – classificação, regressão, agrupamento etc. – para criar uma regra que liga uma entrada a uma saída, usando exemplos já conhecidos. Portanto, esse processo de aprendizado envolve comparar o resultado calculado com o resultado previsto, ajustando o erro computado para alcançar o produto desejado, conforme descrito por Pugliese, Regondi e Marini (2021).

3.3.1 K-Nearest Neighbors (KNN)

Como apresentado por Goldberger *et al.* (2004), o algoritmo K-Nearest Neighbors (KNN) é notavelmente simples, entretanto, surpreendentemente eficaz quando utilizado como um método de classificação. Uma das principais características do KNN é sua capacidade de formar superfícies de decisão não lineares. Isso significa que, ao contrário de alguns métodos de classificação que só conseguem criar linhas retas ou superfícies planas para separar diferentes categorias (ou classes) de dados, o KNN pode criar fronteiras de separação que têm formas irregulares e complexas. Essas formas irregulares são cruciais porque muitos conjuntos de dados do mundo real não são perfeitamente separados por linhas retas.

O algoritmo KNN funciona com base na ideia de proximidade entre os dados. Para classificar um novo ponto, ele considera os " k " pontos mais próximos já conhecidos, sendo " k " um número inteiro previamente definido. A escolha do valor de " k " é crucial para o desempenho do algoritmo: se " k " for muito pequeno, o modelo pode se tornar sensível a ruídos nos dados, classificando de forma incorreta devido à influência de *outliers* (valores com diferença drástica de todos os outros). Por outro lado, um valor de " k " muito grande pode fazer com que o modelo perca a capacidade de capturar padrões mais sutis, resultando em classificações menos precisas.

Outra característica importante do KNN é que sua performance tende a melhorar com o aumento do volume de dados de treinamento. À medida que mais dados são fornecidos, o

algoritmo tem mais exemplos para considerar ao fazer suas classificações, o que geralmente resulta em previsões mais precisas. Isso ocorre porque a presença de mais dados reduz a influência de *outliers* e ruídos, permitindo ao algoritmo identificar padrões mais consistentes e generalizáveis.

3.3.2 Regressão Logística

A Regressão Logística é um método amplamente utilizado para modelagem preditiva em situações em que a variável dependente é binária. Apesar do algoritmo ser chamado de "regressão", é, na verdade, um modelo de classificação. Em algumas literaturas, este modelo é conhecido por outros nomes como classificador de Entropia Máxima (MaxEnt) ou classificador Log-Linear.

O funcionamento do modelo baseia-se na utilização da função logística, também conhecida como sigmoide, para obter uma probabilidade que varia entre 0 e 1. A equação que expressa este comportamento é apresentada na Equação 4 (Berkson, 1944):

$$\hat{Q} = \frac{1}{1 + e^{-(a + bx)}} \quad (4)$$

- x é a variável independente,
- a é o termo de interceptação, ou seja, o valor de y quando x é zero,
- b é o coeficiente de regressão, que multiplica a variável independente x .

3.3.3 Naive-Bayes

O algoritmo de classificação de Naive-Bayes opera sob o princípio do teorema de Bayes com a suposição de independência condicional entre as características. De acordo com Zhang (2004), esse teorema apresenta como premissa fundamental a Equação 5:

$$P(C_k | x_1, \dots, x_n) = \frac{P(C_k) \prod_{i=1}^n P(x_i | C_k)}{P(x_1, \dots, x_n)} \quad (5)$$

- C_k é a variável que representa a classe à qual a instância pode pertencer,
- x_1, \dots, x_n são os atributos observados da instância que está sendo classificada.

Conforme destacado por Rish (2001), apesar de basear-se na hipótese simplificadora de que as características são independentes dado o rótulo da classe, Naive-Bayes demonstra ser notavelmente bem-sucedido na prática, chegando a competir com técnicas mais sofisticadas. Esse fenômeno pode ser explicado pelo fato de que a capacidade de um classificador ser ótimo não depende necessariamente da precisão com que o modelo se ajusta à distribuição de probabilidade real dos dados, ou seja, não é prejudicada pela suposição de independência. O importante é que as distribuições real e estimada concordem quanto à classe mais provável. Isso significa que, mesmo com a independência assumida, Naive-Bayes pode identificar corretamente a classe mais provável e, assim, funcionar de forma eficaz.

3.3.4 Aplicações de aprendizado de máquina em análises no infravermelho

A aplicação de técnicas de aprendizado de máquina em análises de espectros de infravermelho se mostra uma abordagem promissora para a identificação e caracterização de compostos químicos. A interpretação e análise manual motivou o desenvolvimento de ferramentas automatizadas que utilizam aprendizado de máquina para facilitar esse processo.

Uma das aplicações em análises de FT-IR é a identificação de grupos funcionais específicos. Técnicas como redes neurais convolucionais (CNNs) têm sido implementadas para reconhecer padrões espectrais associados a diferentes grupos funcionais. Por exemplo, Enders *et al.* (2021) desenvolveram um modelo de CNNs para identificar grupos funcionais em espectros de FT-IR, utilizando uma base de dados extensa de espectros de compostos orgânicos. Esse modelo demonstrou alta precisão na identificação de grupos funcionais, facilitando a análise de amostras complexas.

4 METODOLOGIA

Neste item, serão detalhados os procedimentos e técnicas utilizados para o desenvolvimento deste trabalho. Serão abordados todos os aspectos metodológicos, desde a coleta de dados espectrais até a implementação dos algoritmos de classificação e aprendizado de máquina.

Primeiramente, será descrito o processo de extração e pré-processamento dos espectros de FT-IR provenientes do banco de dados SDBS. Em seguida, a construção de um banco de dados SQL será detalhada, destacando seu papel no armazenamento e recuperação eficiente dos dados processados. Também será explicada a aplicação dos algoritmos de classificação, incluindo técnicas estatísticas e de aprendizado de máquina, para a análise dos espectros. Por fim, serão apresentados os métodos de avaliação dos resultados, com destaque para as métricas utilizadas na validação da eficácia dos algoritmos propostos.

4.1 COLETA, PRÉ-TRATAMENTO DOS ESPECTROS E EXTRAÇÃO DOS DADOS NUMÉRICOS

As etapas de coleta e o tratamento dos dados, descritas detalhadamente nessa seção, são de grande importância para o funcionamento pleno do programa, pois somente com dados de qualidade é possível obter resultados de qualidade. A seguinte seção entrará em detalhes sobre a fonte dos espectros, o procedimento de coleta e o pré-processamento dos dados para posterior armazenamento.

4.1.1 Descrição da fonte dos dados

A coleta dos espectros IV foi realizada utilizando a base de dados *Spectral Database for Organic Compounds* (SDBS)¹, mantida pelo *National Institute of Advanced Industrial Science and Technology* (AIST) do Japão. Este banco de dados de acesso livre inclui espectros de diversas técnicas analíticas, como FT-IR, ¹H-NMR, ¹³C-NMR, EI-MS, Raman e EPR. A Figura 1, apresentada na Seção 3.1.2 deste trabalho, ilustra o formato original em que o espectro é obtido quando baixado diretamente desse banco de dados.

¹ <https://sdb.db.aist.go.jp>

4.1.2 Procedimentos de coleta

Para garantir a relevância dos dados, fez-se uma revisão bibliográfica do livro de espectros de pesticidas de Visser (2020), onde todos os espectros de compostos com alta resolução, cobrindo a faixa de 4000 a 400 cm^{-1} para FT-IR, que estivessem presentes tanto no banco de dados SDBS quanto no livro, foram coletados. Os espectros provenientes do SDBS foram utilizados para compor o banco de dados, enquanto os espectros do livro foram selecionados para serem utilizados posteriormente como teste dos algoritmos de classificação.

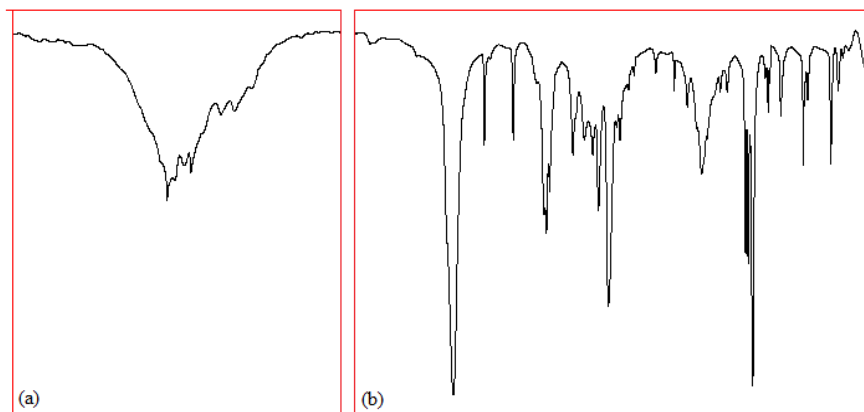
4.1.3 Pré-processamento dos espectros IV

Os espectros IV coletados passaram por um processo de pré-tratamento para garantir a qualidade e comparabilidade dos dados para as etapas seguintes de extração dos dados. As etapas de pré-tratamento incluíram: Recorte da imagem original e remoção de ruídos.

Após a coleta do espectro em seu estado original, este é submetido a um processo no qual é dividido em duas imagens distintas. A primeira seção abrange as bandas de 4000 cm^{-1} até 2000 cm^{-1} , e a segunda, de 2000 cm^{-1} até 400 cm^{-1} . Essa divisão é necessária porque, geralmente, os espectros são representados graficamente com escalas diferentes no eixo dos números de onda, esse procedimento faz com que a extração dos dados quantitativos siga a escala correta para todo o espectro IV.

Após o processo de recorte, elimina-se de forma automática, seguindo o formato padrão do espectro, qualquer ruído visual como manchas, anotações, ou detalhes visuais fora das bandas. Por fim, cada seção é delineada com um contorno vermelho para destacar o tracejado preto. A Figura 4 exemplifica este tratamento ao mostrar o espectro IV do Ácido 1-naftalenoacético no formato final.

Figura 4 – Espectro IV do Ácido 1-naftalenoacético dividido em (a) 4000 – 2000 cm^{-1} e (b) 2000 – 400 cm^{-1} devidamente contornado.



Fonte: Spectral Database for Organic Compounds (SDBS). **Espectro de Infravermelho; SDBS N°: 7054; IR-NIDA-53800**. Disponível em: <https://sdb.sdb.aist.go.jp/>. Acesso em: 30 jun. 2024.

4.1.4 Conversão do espectro IV de imagem para dados numéricos

Para possibilitar a comparação dos dados dos espectros, é necessário que estes estejam em formato numérico. Nesta seção detalha-se o processo de extração dos dados quantitativos a partir das imagens dos espectros tratados, conforme descrito na Seção 4.1.3. Em seguida, explica-se a metodologia de correção de linha de base utilizada, aplicada tanto no banco de dados quanto nos espectros de entrada do programa.

4.1.4.1 Extração dos dados quantitativos

As imagens após o processamento são empregadas em um *script* de automação em batelada feito em JavaScript, que utiliza a ferramenta WebPlotDigitizer² para extrair quantitativamente os dados espectrais com base em parâmetros definidos de número de onda – 4000 até 400 cm^{-1} – e transmitância – 0 até 100%. Ao longo do trabalho, foram utilizados tamanho de malha com o padrão de 1 pixel para Δx e Δy para todas as imagens.

A coleta, o processamento e a extração dos dados quantitativos dos espectros IV são conduzidos separadamente da ferramenta interativa, que é o principal foco deste trabalho. Entretanto, os dados coletados são essenciais para alimentar os algoritmos responsáveis pela

² ROHATGI, A. *WebPlotDigitizer*: Versão 4.7, 2024. Disponível em: <<https://automeris.io/WebPlotDigitizer>>. Acesso em 03 de jul. 2024.

classificação e análise. Na Tabela 1 observa-se um resumo dos dados quantitativos obtidos do espectro IV do Ácido 1-naftalenoacético.

Tabela 1 – Resumo da estrutura dos dados do espectro IV do Ácido 1-naftalenoacético após uso do WebPlotDigitizer.

Número de onda (cm ⁻¹)	Transmitância (%)
410.862	91.5799
409.757	92.1449
...	...
3967.13	76.2381
3926.98	76.5273

Fonte: Autor.

4.1.4.2 Correção da linha de base

Os sinais de instrumentos analíticos consistem em informações químicas, linha de base e ruídos aleatórios. A presença desses ruídos sobre a linha de base pode impactar negativamente tanto em resultados analíticos quantitativos quanto qualitativos, sendo essencial que essa influência seja mitigada (Zhang; Chen; Liang, 2010). Neste contexto, Zhang, Chen e Liang propuseram e desenvolveram um algoritmo rápido e flexível para o ajuste da linha de base, baseado no método de quadrados mínimos penalizados ponderados iterativamente e adaptativos (*Adaptive iteratively reweighted penalized least squares* – airPLS³).

Para a padronização dos espectros, aplica-se a versão em Python do algoritmo de airPLS, em conjunto com o algoritmo de suavização de Whittaker, também conhecido como algoritmo de quadrados mínimos penalizados. Este último que serviu como base para a implementação completa do airPLS.

O airPLS utiliza três variáveis de entrada principais: “x” sendo o espectro a ser processado, um parâmetro “*lambda_*” que controla a intensidade da suavização da linha de base, “*porder*” que determina a ordem da diferença penalizada – permitindo ajustes mais finos para curvaturas mais suaves ou mudanças graduais – e “*itermax*”, que define o número máximo de iterações permitidas no processo de ajuste da linha de base.

³ Z.-M. ZHANG, S. CHEN, Y.-Z. LIANG. *airPLS*: Versão 2.0, 2016. Disponível em: <<https://github.com/zmzhang/airPLS>>. Acesso em 03 de jul. 2024.

4.2 DESENVOLVIMENTO DO BANCO DE DADOS SQL

Os dados quantitativos derivados dos espectros IV, coletados conforme o método descrito na Seção 4.1 são armazenados em um banco de dados SQL, utilizando um arquivo “.db” provido pela biblioteca SQLite. O desenvolvimento de um banco de dados SQL é essencial para o gerenciamento eficiente dos dados espectrais coletados. Ele permite a organização sistemática dos dados, facilitando a recuperação rápida e precisa das informações necessárias para a análise e classificação dos espectros.

No entanto, a maneira como esses dados são organizados para utilização no algoritmo desenvolvido e nos modelos de aprendizado de máquina apresenta diferenças significativas, resultando em estruturas de armazenamento distintas. Isso ocorre porque o algoritmo desenvolvido e os algoritmos de aprendizado de máquina manipulam os dados de formas diferentes. Essas abordagens serão detalhadas nas seções a seguir.

4.2.1 Configuração para o algoritmo principal

Para o algoritmo central deste estudo, não foram necessárias modificações adicionais nos valores já ilustrados na Tabela 1. Utilizando SQL, são estruturadas duas tabelas para organização dos dados: a primeira tabela contém o nome do composto e um número de identificação único para cada um; a segunda tabela armazena os dados de número de onda e transmitância, que estão vinculados ao número de identificação único referenciado na primeira tabela.

4.2.2 Configuração para os algoritmos de aprendizado de máquina

Para o processamento dos dados nos modelos de aprendizado de máquina, os números de onda são primeiramente segmentados em intervalos. Foram investigados intervalos variando entre 1 cm^{-1} e 100 cm^{-1} . No entanto, o intervalo que apresentou os melhores resultados foi o de 25 cm^{-1} . Dentro de cada intervalo, o valor máximo de transmitância é selecionado para representar a faixa. Essa abordagem resulta em uma linha completa de dados correspondente a um único composto.

A escolha do intervalo de 25 cm^{-1} como a segmentação mais eficaz baseia-se na precisão das classificações obtidas. Essa precisão se refere à capacidade do modelo de identificar os compostos corretos com maior acurácia. Ou seja, ao utilizar intervalos de 25 cm^{-1} , o modelo

foi capaz de definir maior similaridade entre os compostos de entrada e do banco de dados de maneira correta.

Cada composto no banco de dados foi tratado de maneira idêntica, permitindo que os modelos de aprendizado de máquina utilizem uma análise sequencial para interpretação dos dados. Essencialmente, o modelo aprende a reconhecer padrões e fazer indicações baseadas na similaridade dos valores de transmitância em colunas correspondentes entre diferentes compostos. A Tabela 2 oferece um exemplo simplificado da tabela de dados usada para treinar os algoritmos na classificação dos espectros de interesse.

Tabela 2 – Estrutura exemplo resumida utilizada para a análise com os algoritmos de aprendizado de máquina.

Nome	400 – 425	425 – 450	...	3925 – 3950	3950 – 3975	3975 – 4000
1,2-Dichloropropane	0.272	0	...	0	0.484	0.536
1,3-Dichloropropene(cis)	0	0.779	...	0.448	0.301	0
...
Triforine	0.287	0	...	0	0	0.327
Vinclozolin	0	0.286	...	0	0.220	0

Fonte: Autor.

4.3 ARQUITETURA E FUNCIONAMENTO DO ALGORITMO

Este subtópico descreve a estrutura do algoritmo, explicando como ele foi projetado para processar os dados espectrais, realizar a classificação e produzir os resultados. Serão abordados os componentes principais do algoritmo, incluindo os métodos de pré-processamento e a integração com o banco de dados SQL. Além disso, serão apresentados os critérios de avaliação e validação do algoritmo, demonstrando como ele foi ajustado para otimizar seu desempenho.

4.3.1 Filtragem das bandas de interesse

A metodologia implementada para filtrar bandas de interesse dos espectros IV visa otimizar a acurácia da classificação sem sobrecarregar o algoritmo com uma quantidade excessiva de dados. As etapas que compõem esse processo são a normalização da transmitância e a estratégia de seleção de bandas relevantes para a identificação de compostos.

Inicialmente, realiza-se o ajuste da linha de base como descrito na Seção 4.1.4.2 para os dados de entrada, a normalização dos dados de transmitância é o passo seguinte. Neste processo, o ponto de maior transmitância dentro do espectro é utilizado como referência, subtraindo-se seu valor da transmitância observada em cada ponto ao longo do eixo de número de onda do espectro. Essa abordagem resulta em uma série de valores que representam as alturas relativas das bandas.

Para a seleção efetiva das bandas, implementa-se uma metodologia que envolve a busca por mínimos locais, considerando uma janela de análise, ou uma faixa de valores, específica. No contexto deste trabalho, assim como feito na Seção 4.2.2, foram investigadas faixas variando entre 10 cm^{-1} e 40 cm^{-1} . Essas faixas estão disponíveis para ajuste na interface da aplicação. A faixa padrão que apresentou os melhores resultados foi a de 25 cm^{-1} . O algoritmo analisa tanto 25 cm^{-1} antes quanto 25 cm^{-1} após cada ponto no eixo do número de onda. Durante essa varredura, compara-se a altura de cada banda com as alturas nas proximidades dentro da faixa especificada.

Uma banda é considerada de alta relevância se, ao longo dessa faixa de 25 cm^{-1} em ambas as direções, a altura relativa observada se mantém como a mais elevada, indicando um pico proeminente. Essas bandas selecionadas são classificadas como essenciais para a etapa subsequente de cálculo de similaridade.

4.3.2 Algoritmo de similaridade

O algoritmo desenvolvido para a análise de similaridade entre espectros inicia-se com uma função de comparação e filtragem dos dados espectrais. Este processo envolve a identificação das melhores correspondências entre os números de onda do espectro de entrada e do espectro da base de dados de componentes.

Cada banda do espectro de entrada é comparada com as bandas do espectro da base de dados, buscando minimizar as diferenças entre suas posições (números de onda). Este passo é crucial para assegurar que as bandas comparadas apresentem a possibilidade de serem dos mesmos componentes químicos nos diferentes espectros, ou seja, representam o mesmo grupo funcional. Caso os números de onda não correspondam – indicando uma discrepância significativa – essa banda específica é considerada não comparável e é pareada com um valor zero, gerando impacto no cálculo de similaridade.

Após estabelecer as melhores correspondências, o algoritmo procede ao cálculo da similaridade, adotando pesos diferenciados para cada banda espectral. Esses pesos são baseados

na altura relativa de cada banda em relação à linha de base do espectro. Nesta metodologia, bandas que ultrapassam 50% da altura relativa da linha de base são consideradas "grandes" e recebem um peso de dois, enquanto bandas acima de 30% são classificadas como "médias" e recebem um peso de um. Bandas "pequenas" abaixo de 30% são desconsideradas.

Segundo Da Silva *et al.* (2016), o coeficiente de correlação de Pearson é um parâmetro conveniente para a análise de similaridade entre espectros. Este coeficiente, que varia de -1 a +1, indica o grau de correlação linear entre dois conjuntos de dados. Um coeficiente de +1 sugere uma relação diretamente proporcional, enquanto -1 indica uma relação inversamente proporcional. Valores próximos de zero sugerem ausência de correlação linear. Para facilitar a interpretação nas comparações espectrais, este coeficiente foi adaptado para expressar porcentagem de similaridade. Assim, correlações altas, entre 100% e 80%, indicam grande similaridade, e correlações médias, entre 80% e 50%, sugerem similaridade moderada. A implementação da correlação de Pearson foi realizada utilizando a biblioteca Scipy⁴.

O algoritmo calcula o coeficiente de correlação de Pearson para três parâmetros diferentes: número de onda, transmitância e altura das bandas. A média desses três coeficientes é o valor resultante de saída para o usuário.

4.4 BIBLIOTECAS DE ALGORITMOS DE APRENDIZADO DE MÁQUINA

Neste tópico, serão abordadas as técnicas de normalização de dados e as bibliotecas de algoritmos de aprendizado de máquina utilizadas no desenvolvimento deste trabalho, levando em consideração os parâmetros de configuração de cada modelo.

4.4.1 Normalização dos dados

Os dados dos espectros obtidos na Seção 4.2.2 são normalizados utilizando a classe de pré-processamento "MinMaxScaler" da biblioteca Scikit-Learn⁵. Esta técnica é essencial para ajustar os valores do conjunto de dados a uma escala comum, especificamente entre zero e um. A classe é então aplicada em todas as colunas relativas às faixas de número de onda individualmente.

A normalização é crucial porque garante que as variáveis que possuem escalas diferentes não influenciem de forma desproporcional o modelo de análise. Além disso, esse

⁴ <https://scipy.org/>

⁵ <https://scikit-learn.org/stable/>

processo pode ajudar a melhorar a convergência dos algoritmos de aprendizado de máquina e aumentar a sua precisão.

4.4.2 Modelos de classificação por aprendizado de máquina

Os modelos de aprendizado de máquina empregados para a classificação dos espectros de interesse incluíram o K-Nearest Neighbors (KNN), Regressão Logística e Naive Bayes. Esses algoritmos foram implementados utilizando a biblioteca Scikit-Learn em Python.

A Regressão Logística foi aplicada em sua forma convencional, onde o conjunto de dados é dividido em dados de treinamento e de teste. Este modelo calcula a probabilidade de associação às classes previstas com base nas características dos dados fornecidos.

Por outro lado, os modelos KNN e Naive Bayes foram manipulados de maneira distinta. Ambos os métodos identificam inicialmente o composto mais similar ao espectro em análise. Para obter os cinco compostos mais relevantes, foi necessário implementar um loop que se repete cinco vezes. Em cada iteração, o composto previamente identificado como o mais similar é removido da lista de candidatos, permitindo que o algoritmo identifique o próximo composto mais similar sem repetições. Este processo assegura a obtenção de uma lista dos cinco principais compostos correlacionados ao espectro analisado.

Usualmente, algoritmos de aprendizado de máquina possuem atributos chamados hiperparâmetros. Estes são variáveis de configuração externa utilizadas para controlar e gerenciar o treinamento do modelo. Os hiperparâmetros utilizados nos algoritmos foram os seguintes: para KNN, utilizou-se $k=3$; para Regressão Logística, foi configurado $max_iter=1000$; e para Naive-Bayes, não foi necessário o ajuste de nenhum hiperparâmetro.

5 RESULTADOS E DISCUSSÕES

Neste capítulo, são apresentados e discutidos os resultados obtidos a partir das análises realizadas com o sistema desenvolvido. Serão expostos os dados coletados e o desempenho dos algoritmos de classificação. A discussão incluirá uma comparação entre os resultados obtidos com o algoritmo desenvolvido e os algoritmos de aprendizado de máquina utilizados. Este capítulo tem como objetivo proporcionar uma compreensão aprofundada dos achados do trabalho e sua relevância para o campo de estudo.

5.1 ESTRUTURA DO BANCO DE DADOS

Com a metodologia descrita nas Seções 4.1 e 4.2, foi construído um banco de dados composto por 151 (cento e cinquenta e um) agrotóxicos, obtidos através da base de dados SDBS. A listagem completa desses compostos, que foram empregados como referências na aplicação, está documentada no Apêndice A.

5.2 CRITÉRIOS DE SELEÇÃO DAS AMOSTRAS PARA ANÁLISE

As amostras para análise foram selecionadas aleatoriamente do banco de dados descrito na Seção 5.1. Utilizando um método de escolha aleatória com o auxílio de Python, dez compostos foram extraídos do total de cento e cinquenta e um derivados. A escolha foi feita diretamente do banco de dados contendo todos os espectros, selecionando índices aleatórios representativos de cada composto. Posteriormente, foi realizada uma verificação para substituir compostos com geometria molecular muito semelhante, como, por exemplo, o Triadimefom e o Triadomenol. No entanto, evitou-se especificidade excessiva ou interferência direta na escolha dos espectros, a fim de prevenir a alienação dos algoritmos de aprendizado de máquina, que poderia resultar em *overfitting*. O *overfitting* ocorre quando os resultados dos testes se encaixam perfeitamente nos dados de treinamento, mas não são preditivos para casos reais. Esse tipo de problema não ocorre no algoritmo desenvolvido.

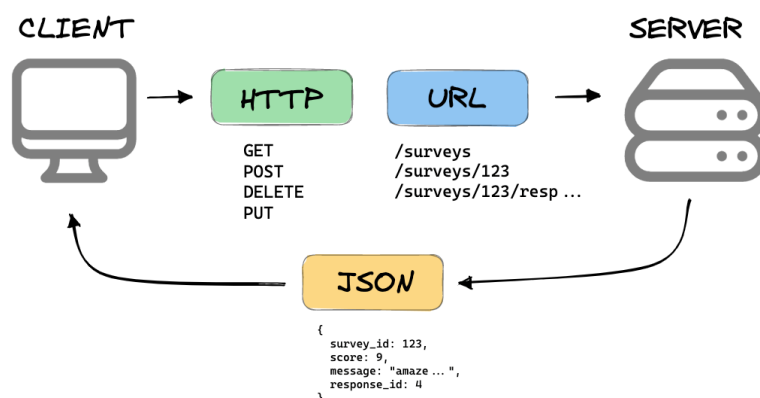
Os espectros correspondentes a esses dez compostos foram obtidos de Visser (2020), seguindo a metodologia detalhada na Seção 4.1. Os compostos selecionados para compor o conjunto de dados de entrada incluem: Ácido 1-naftalenoacético, Azociclotina, Bendiocarbe, Binapacril, Coumafós, Sulfato de endossulfam, Fention, Lenacil, Fosadona e Triadimefom.

5.3 INTERFACE DE USUÁRIO E INTEGRAÇÃO DA ARQUITETURA

A interface de usuário do aplicativo foi projetada utilizando a ferramenta Vite, em conjunto com a biblioteca React.js e CSS para oferecer uma experiência intuitiva e fácil de navegar, caracterizada por um design minimalista. O software é composto por duas abas principais: “*Home*” (Entrada de Resultados) e “Relatórios”.

Os algoritmos de *backend* foram desenvolvidos em Python, utilizando a arquitetura API Rest, com suporte do *framework* FastAPI. Esta arquitetura é estruturada em torno de terminais de Interface de Programação de Aplicação (*Application Programming Interface* – API), que contém o Identificador Uniforme de Recursos (*Uniform Resource Identifier* – URI) – neste caso, os algoritmos. As operações sobre esses recursos são realizadas através de métodos de HTTP, como POST ou GET. Os resultados dos algoritmos são enviados de volta ao cliente em formato JSON, permitindo que sejam facilmente acessados e manipulados na interface do usuário. A Figura 5 apresenta um fluxograma simplificado da metodologia API Rest, ilustrando a interação entre o cliente e o servidor.

Figura 5 – Fluxograma simplificado da metodologia API Rest.



Fonte: Mann, H. "REST API." Mannhowie.com. Disponível em: <<https://mannhowie.com/rest-api>>. Acesso em 03 de jul. 2024.

A aba “*Home*” foi organizada para incluir tanto funções simples quanto avançadas, facilitando a experiência do usuário conforme sua familiaridade com a ferramenta. Na seção de configurações básicas, o usuário precisa apenas importar o arquivo do espectro IV em formato ".csv" e nomear a amostra. Já nas configurações avançadas, é possível otimizar o processamento ao selecionar o uso de múltiplos núcleos do processador, definir a faixa de análise conforme

descrito na Seção 4.3.1, e ajustar os parâmetros do algoritmo de correção de linha de base (airPLS), detalhados na Seção 4.1.4.2.

Caso prefira-se uma abordagem direta, há a opção de manter as configurações avançadas como estão e iniciar a análise clicando no botão "Analisar Composto", que direciona o usuário para a aba de relatórios. A Figura 6 apresenta uma captura de tela da interface inicial.

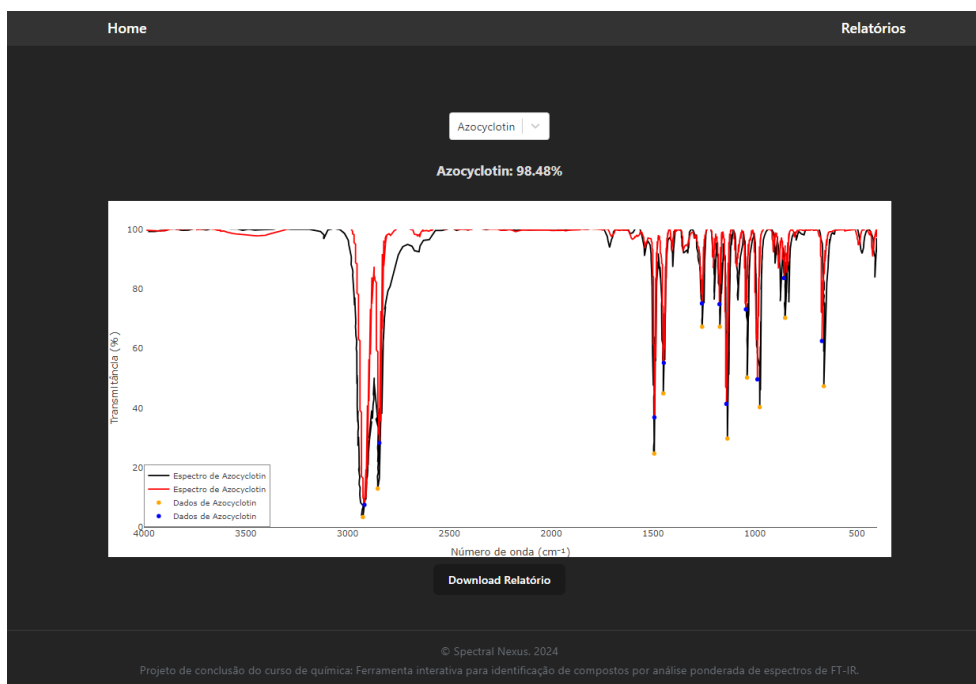
Figura 6 – Interface interativa da tela inicial do usuário (*Home*).

A interface da tela inicial do usuário, intitulada "Home", apresenta uma barra de navegação superior com "Home" e "Relatórios". O formulário principal é dividido em seções: "Configurações:" com um botão "Escolher arquivo" e o nome do arquivo "Azocyclotin.csv" exibido; "Configurações Avançadas:" com uma opção "Usar paralelização: On" marcada, um seletor de "Núcleos" com o valor "16", uma barra deslizante para "Faixa de análise (10 - 40)" com o valor "25", e campos de entrada para "Lambda:" (100), "P order:" (1) e "Max iterações:" (15). Um botão "Analisar composto" está centralizado na base do formulário. O rodapé contém o copyright "© Spectral Nexus. 2024" e o texto "Projeto de conclusão do curso de química: Ferramenta interativa para identificação de compostos por análise ponderada de espectros de FT-IR."

Fonte: Autor.

Na aba "Relatórios", o usuário pode explorar uma lista dos cinco compostos que apresentaram maior similaridade conforme determinado pelo algoritmo. Ao selecionar qualquer um destes compostos na lista, será exibido um gráfico detalhado mostrando a sobreposição dos espectros IV. Acima do gráfico, é apresentado a porcentagem indicando o grau de similaridade entre os espectros. Além disso, o usuário tem a opção de baixar um relatório completo, que inclui todas as informações detalhadas desta aba, clicando no botão "Download Relatório". Uma captura de tela da interface de relatórios está disposta na Figura 7.

Figura 7 – Interface interativa da tela de relatórios.



Fonte: Autor.

O código-fonte desta aplicação pode ser acessado e modificado livremente através do repositório spectral-nexus⁶ no GitHub.

5.4 CONCEPÇÃO E DESIGN DO RELATÓRIO DE SAÍDA

O relatório gerado pela aplicação foi projetado para auxiliar o usuário na identificação de possíveis discrepâncias ou imprecisões no algoritmo. Ao listar os cinco compostos mais similares encontrados no banco de dados, este proporciona ao usuário uma variedade de opções para análise e seleção do que melhor corresponde à sua amostra. O design utilizado permite uma avaliação mais refinada e uma escolha mais informada por parte do operador.

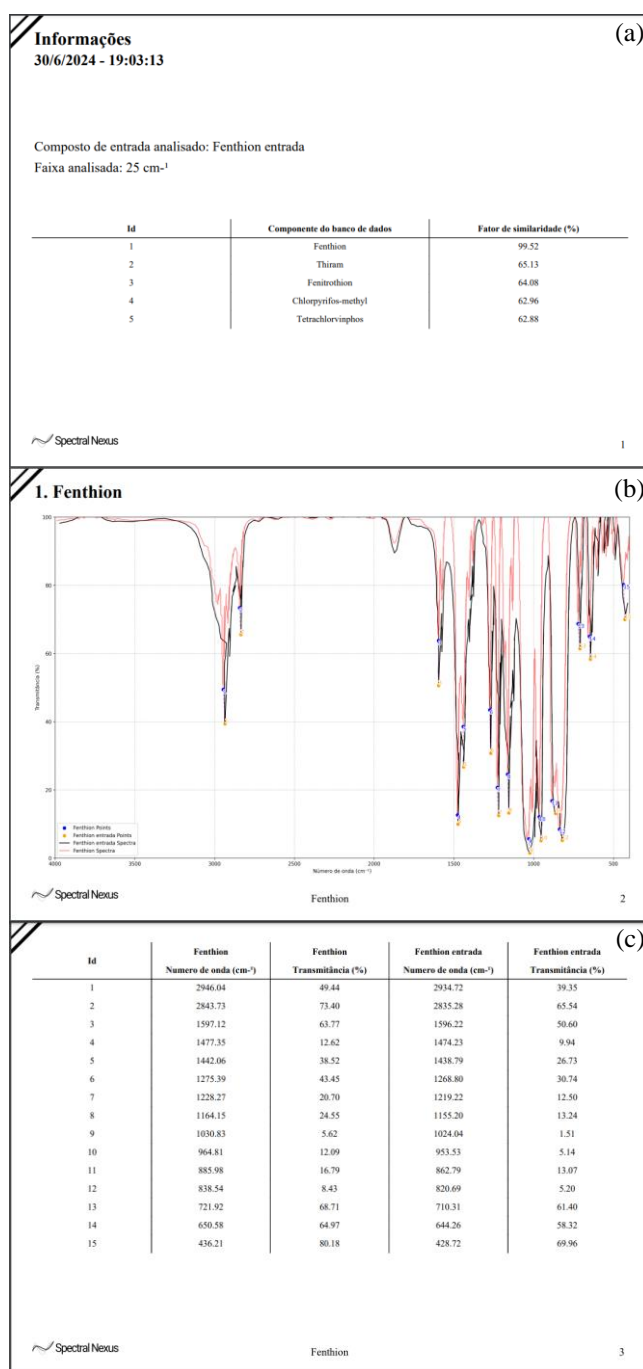
Na Figura 8, ilustra-se um exemplo das primeiras páginas do relatório de saída. A primeira página detalha informações como a data e a hora de geração do relatório, o nome atribuído ao composto analisado, a faixa espectral em número de onda utilizada para a varredura e uma tabela que classifica os compostos do banco de dados por ordem de similaridade, incluindo os respectivos fatores de similaridade ao lado de cada entrada.

As páginas subsequentes exibem o espectro IV obtido do composto de referência do banco de dados em comparação com o espectro IV da amostra do usuário. Estes espectros são

⁶ Motta, J. G. (2024). *Spectral Nexus*. GitHub. Disponível em: <<https://github.com/jgmotta98/spectral-nexus>>.

complementados por pontos de interesse marcados, que representam as bandas comparadas entre a referência e a amostra. Alternando com esses espectros, tabelas detalham os dados quantitativos relacionados às bandas de interesse. Este formato de apresentação com espectros e tabelas correspondentes continua até que os cinco compostos mais similares tenham sido completamente apresentados.

Figura 8 – (a) Informações resumidas sobre a análise do composto. (b) Espectros IV. (c) Tabela contendo as bandas de interesse. As páginas seguintes repetem o padrão de (b) e (c) para os outros componentes.



Fonte: Autor.

5.5 ANÁLISE DAS AMOSTRAS

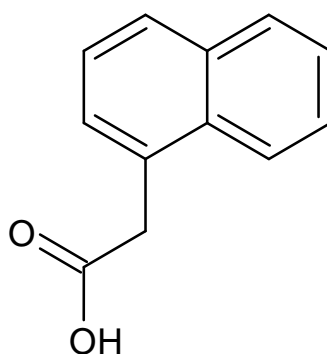
Neste subtópico, apresentados os resultados da análise, destacando as principais características observadas nos espectros e a eficácia dos algoritmos de classificação na identificação dos compostos. Esta análise é fundamental para validar a robustez e a precisão do sistema desenvolvido. Como descrito na Seção 5.2, os compostos selecionados de Visser (2020) para serem utilizados como amostra foram: Ácido 1-naftalenoacético, Azociclotina, Bendiocarbe, Binapacril, Coumafós, Sulfato de endossulfam, Fention, Lenacil, Fosadona e Triadimefom.

Notou-se durante a análise que diferenças significativas na porcentagem de similaridade ainda podem representar moléculas estruturalmente similares. Isso ocorre devido à presença de grupos funcionais comuns ou padrões de absorção similares em diferentes compostos. Por exemplo, duas moléculas podem apresentar uma similaridade espectral de apenas 60%, mas ainda compartilhar estruturas químicas essenciais, como anéis aromáticos ou cadeias alifáticas, que influenciam seus espectros de forma semelhante. Portanto, é fundamental considerar não apenas a porcentagem de similaridade, mas também a análise qualitativa das características espectrais e o contexto dos resultados.

5.5.1 Ácido 1-naftalenoacético

O Ácido 1-naftalenoacético é utilizado como regulador de crescimento e atua como pesticida. De acordo com as normativas da União Europeia, ele possui uma ingestão diária aceitável (ADI) estabelecida de 0,1 mg por kg de peso corporal por dia (PubChem, 2024a). Sua estrutura química pode ser observada na Figura 9.

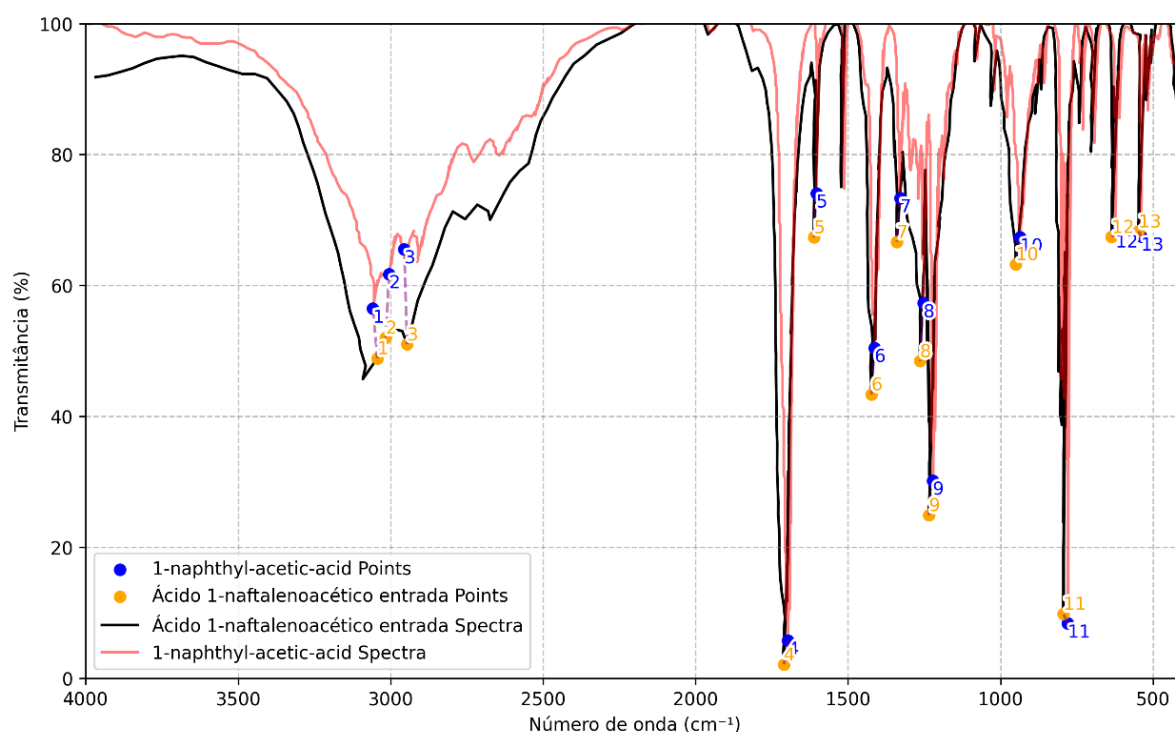
Figura 9 – Representação da estrutura química do Ácido 1-naftalenoacético.



Fonte: PubChem, 2024a.

Após o processamento pelo algoritmo desenvolvido, o relatório gerado, como discutido na Seção 5.4, apresenta uma comparação visual dos espectros IV. Esta visualização detalhada exibe o espectro do composto do banco de dados em sobreposição ao espectro do composto submetido pelo usuário. Tal disposição facilita uma análise das características de cada banda espectral, permitindo ao usuário avaliar a precisão com que o algoritmo seleciona e compara as bandas relevantes. A Figura 10 demonstra essa sobreposição para o Ácido 1-naftalenoacético.

Figura 10 – Espectros IV sobrepostos do Ácido 1-naftalenoacético do banco de dados (vermelho) e de entrada (preto).



Fonte: Autor.

Ainda na Figura 10, observa-se a distribuição de pontos azuis e laranja, que representam as bandas com mínimos locais de interesse identificados pelo algoritmo. A proximidade entre os pontos, tanto em termos de número de onda quanto de transmitância, reflete uma maior correlação de Pearson, que indica a similaridade entre os compostos analisados. A Tabela 3 apresenta os valores detalhados de número de onda e transmitância para cada ponto destacado na Figura 10. As geometrias moleculares de todos os agrotóxicos encontram-se no Apêndice B.

Tabela 3 – Comparação das bandas de interesse entre o Ácido 1-naftalenoacético do banco de dados (Banco de dados) e o de entrada (Amostra).

Índex	Banco de dados Número de onda (cm^{-1})	Amostra Número de onda (cm^{-1})	Banco de dados Transmitância (%)	Amostra Transmitância (%)
1	3058,09	3043,59	56,46	48,90
2	3004,84	3016,04	61,75	52,04
3	2955,40	2946,11	65,55	51,02
4	1697,46	1710,83	5,79	2,10
5	1603,97	1611,88	74,09	67,38
6	1413,17	1423,58	50,48	43,39
7	1329,22	1340,16	73,36	66,68
8	1252,90	1264,31	57,31	48,46
9	1222,37	1234,80	30,21	24,94
10	936,17	950,17	67,37	63,27
11	779,72	794,05	8,39	9,87
12	623,27	635,58	67,97	67,42
13	539,32	547,37	67,50	68,26

Fonte: Autor.

O algoritmo desenvolvido identificou o Ácido 1-naftalenoacético como o composto mais similar, com um fator de similaridade de 63,95%. Este foi seguido por Procimidona (57,09%), MCPA (51,21%), Captafol (48,82%) e Metolacoloro (48,29%). Os métodos de aprendizado de máquina KNN e Naive-Bayes não foram capazes de identificar o Ácido 1-naftalenoacético, entretanto, Regressão Logística, classificou-o como o terceiro composto mais similar. Os resultados de classificação do Ácido 1-naftalenoacético de cada método são visualmente representados na Figura 11.

Figura 11 – Classificação do Ácido 1-naftalenoacético por ordem de similaridade (maior para menor).

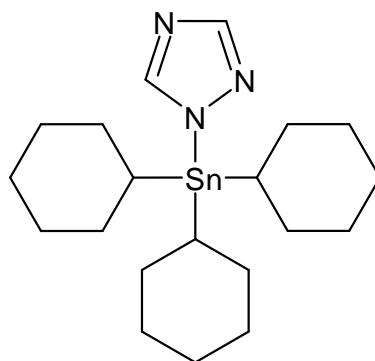
Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
Ácido 1-naftalenoacético	Bendiocarbe	Clorpirifós-metil	Bendiocarbe
Promicidona	Carbofurano	Pirimicarbe	Carbofurano
MCPA	Amitraz	Ácido 1-naftalenoacético	Pirimicarbe
Captafol	Clordimeforme	Clordimeforme	Clordimeforme
Metalacloro	Clorpirifós-metil	Sulfato de endossulfam	Amitraz

Fonte: Autor.

5.5.2 Azociclotina

A Azociclotina, empregada tanto como acaricida quanto pesticida, tem sua ingestão diária aceitável (ADI) definida pelas regulamentações da União Europeia em 0,003 mg por kg de peso corporal diariamente (PubChem, 2024b). A estrutura química deste composto é ilustrada na Figura 12.

Figura 12 – Representação da estrutura química da Azociclotina.

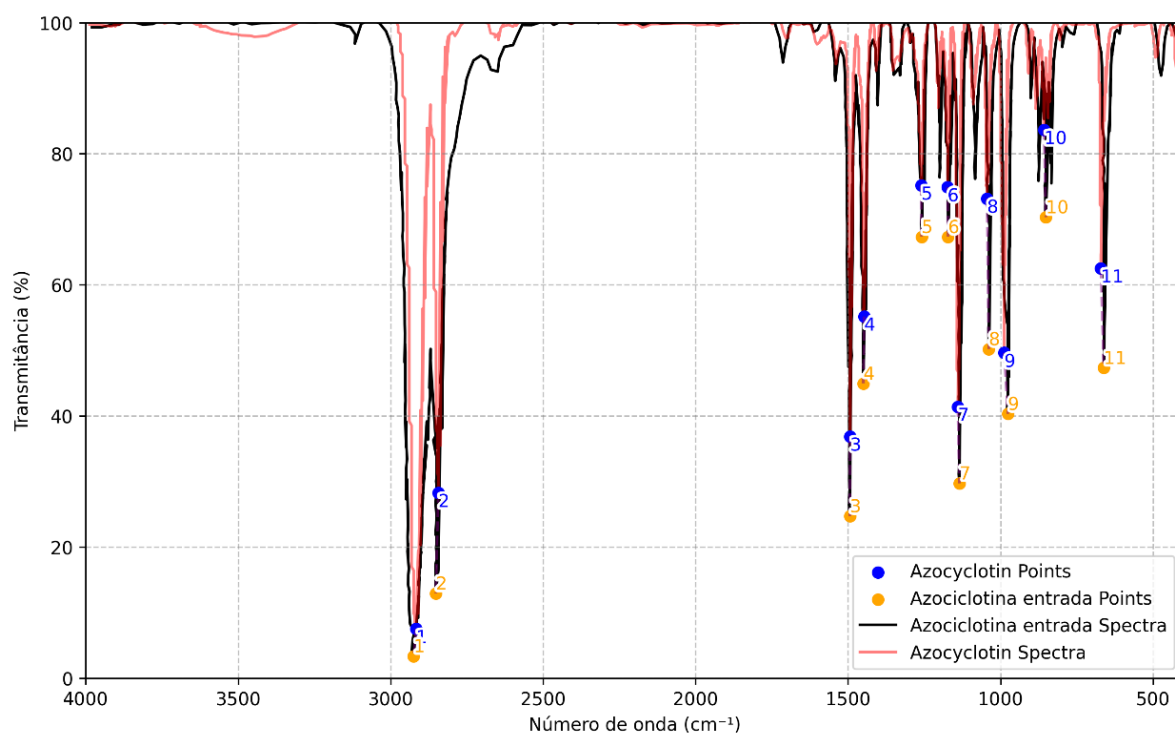


Fonte: Adaptado de PubChem, 2024b.

O relatório resultante, conforme explorado na Seção 5.4 e enfatizado na Seção 5.5.1, ilustra uma comparação dos espectros IV. A sobreposição dos espectros do banco de dados e

de entrada do usuário facilita uma avaliação detalhada das bandas espectrais. A Figura 13 expõe essa comparação para a Azociclotina.

Figura 13 – Espectros IV sobrepostos da Azociclotina do banco de dados (vermelho) e de entrada (preto).



Fonte: Autor.

Como definido anteriormente na Seção 5.5.1, na Tabela 4 seguem os valores especificados de números de onda e transmitância representados por cada um dos pontos azul e laranja da Figura 13.

Tabela 4 – Comparação das bandas de interesse entre a Azociclotina do banco de dados (Banco de dados) e o de entrada (Amostra).

(continua)

Índex	Banco de dados	Amostra	Banco de dados	Amostra
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Transmitância (%)	Transmitância (%)
1	2916,81	2924,43	7,53	3,38
2	2843,73	2851,80	28,27	12,94
3	1493,30	1493,99	36,87	24,73
4	1447,15	1449,67	55,19	44,94
5	1260,00	1258,95	75,17	67,32

(conclusão)				
Índex	Banco de dados	Amostra	Banco de dados	Amostra
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Transmitância (%)	Transmitância (%)
6	1173,98	1171,97	74,91	67,31
7	1139,90	1134,92	41,42	29,74
8	1044,02	1038,48	73,17	50,21
9	987,76	975,93	49,63	40,33
10	857,03	851,43	83,64	70,32
11	670,75	661,48	62,52	47,34

Fonte: Autor.

O algoritmo desenvolvido identificou a Azociclotina como o composto mais similar, atingindo um fator de similaridade de 98,48%. Os compostos subsequentes incluíram Cihexatina (79,29%), Hexaconazol (56,33%), Alloxidium-sodium (52,04%) e Aldicarbe (50,40%). Notavelmente, todos os métodos de aprendizado de máquina empregados — KNN, Regressão Logística e Naive-Bayes — também identificaram a Azociclotina como o composto com maior similaridade. A Figura 14 ilustra os resultados comparativos da classificação da Azociclotina por cada um dos métodos analisados. As estruturas moleculares de todos os pesticidas estão disponíveis no Apêndice C.

Figura 14 – Classificação da Azociclotina por ordem de similaridade (maior para menor).

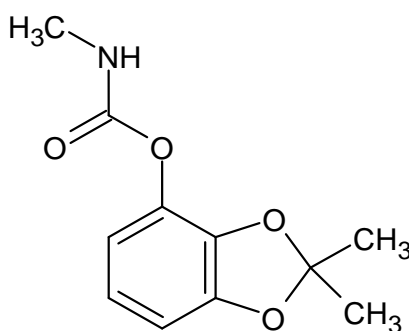
Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
Azociclotina	Azociclotina	Azociclotina	Azociclotina
Cihexatina	Cicloxidime	Dodina	Cicloxidime
Hexaconazol	Carbofurano	Cicloxidime	Dodina
Alloxidium-sodium	Dodina	Trifluralina	Trifluralina
Aldicarbe	Bendiocarbe	Triadimenol	Carbofurano

Fonte: Autor.

5.5.3 Bendiocarbe

O Bendiocarbe é um composto químico utilizado tanto como inseticida quanto pesticida. Segundo as normas da União Europeia, a ingestão diária aceitável (ADI) para este composto é de 0,004 mg por kg de peso corporal por dia (PubChem, 2024c). Sua estrutura química pode ser visualizada na Figura 15.

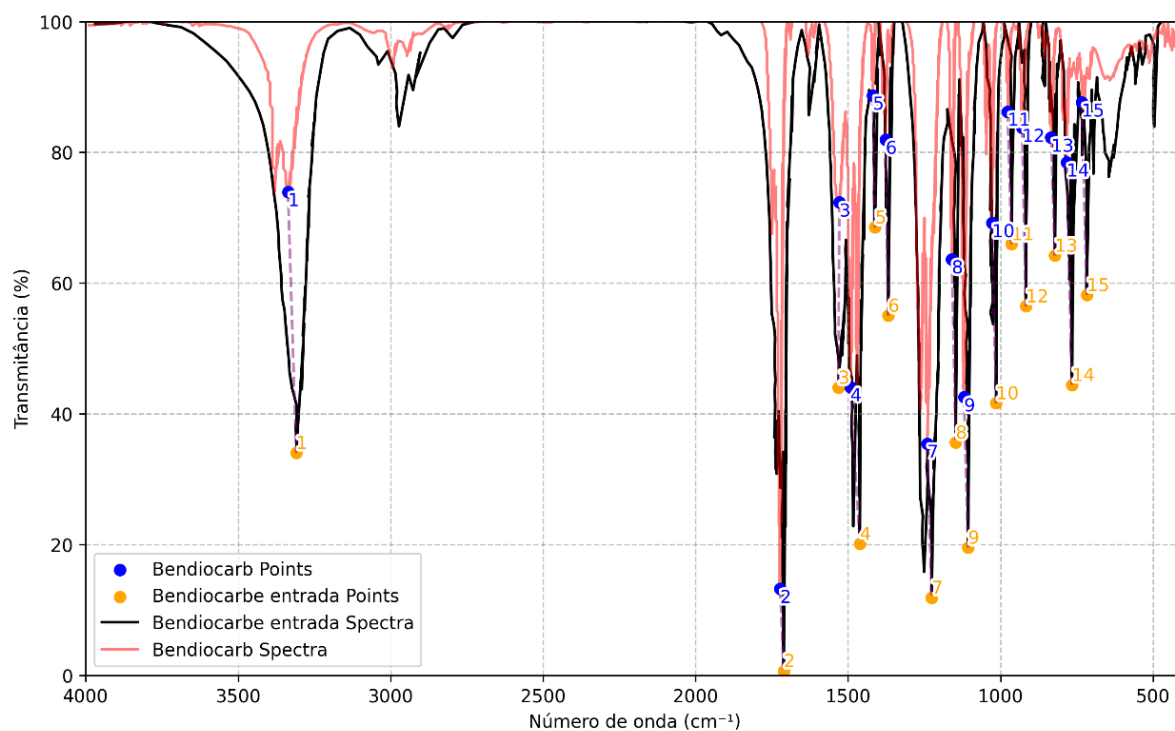
Figura 15 – Representação da estrutura química do Bendiocarbe.



Fonte: Adaptado de PubChem, 2024c.

Conforme detalhado nas seções anteriores, a análise visual no relatório destaca uma sobreposição entre os espectros do composto de referência e o analisado, permitindo uma melhor investigação das bandas selecionadas pelo algoritmo. A Figura 16 demonstra essa sobreposição para o Bendiocarbe.

Figura 16 – Espectros IV sobrepostos do Bendiocarbe do banco de dados (vermelho) e de entrada (preto).



Fonte: Autor.

Como observado anteriormente, os valores especificados de números de onda e transmitância representados por cada um dos pontos azul e laranja na Figura 16 são detalhados na Tabela 5.

Tabela 5 – Comparação das bandas de interesse entre o Bendiocarbe do banco de dados (Banco de dados) e o de entrada (Amostra).

Índex	Banco de dados	Amostra	(continua)	
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Transmitância (%)	Transmitância (%)
1	3336,63	3309,59	73,94	34,05
2	1722,91	1710,73	13,25	0,70
3	1528,59	1531,65	72,33	44,02
4	1492,39	1461,24	44,11	20,14
5	1418,82	1411,35	88,69	68,53
6	1376,87	1368,73	81,93	55,03
7	1240,42	1225,84	35,44	11,90
8	1159,73	1146,86	63,65	35,64

(conclusão)				
Índex	Banco de dados	Amostra	Banco de dados	Amostra
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Transmitância (%)	Transmitância (%)
9	1119,76	1106,46	42,56	19,56
10	1028,10	1014,62	69,19	41,63
11	976,75	964,11	86,23	65,97
12	929,28	916,82	83,76	56,46
13	834,30	821,77	82,25	64,21
14	783,19	765,75	78,45	44,45
15	733,75	717,08	87,63	58,18

Fonte: Autor.

O algoritmo desenvolvido destacou o Bendiocarbe como o composto mais próximo, com um fator de similaridade de 97,17%. Este foi seguido por Cimoxanil (76,94%), Clorobenzilato (74,41%), Desmedifam e Fenmedifam, ambos com 61,74%. Enquanto todos os métodos de aprendizado de máquina reconheceram o Bendiocarbe, apenas KNN o classificou como o mais similar. Por outro lado, Regressão Logística posicionou-o como o quinto mais similar, e Naive-Bayes como o segundo. A Figura 17 apresenta uma comparação visual dos resultados de classificação do Bendiocarbe por cada um desses métodos. Todas as geometrias dos agrotóxicos estão detalhadas no Apêndice D.

Figura 17 – Classificação da Bendiocarbe por ordem de similaridade (maior para menor).

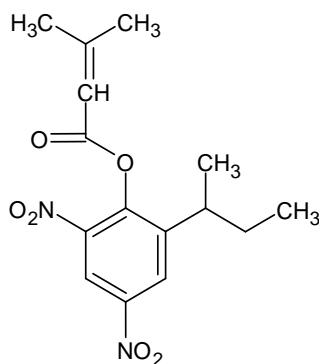
Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
Bendiocarbe	Bendiocarbe	Carbosulfano	Carbofurano
Cimoxanil	Carbofurano	Difenzoquat-metil-sulfato	Bendiocarbe
Clorobenzilato	Cloroprofam	Cloroprofam	Difenzoquat-metil-sulfato
Desmedifam	Desmedifam	Carbofurano	Cloroprofam
Fenmedifam	Difenzoquat-metil-sulfato	Bendiocarbe	Desmedifam

Fonte: Autor.

5.5.4 Binapacril

O Binapacril é um composto químico utilizado tanto como fungicida quanto pesticida (PubChem, 2024d). Sua estrutura química é ilustrada na Figura 18.

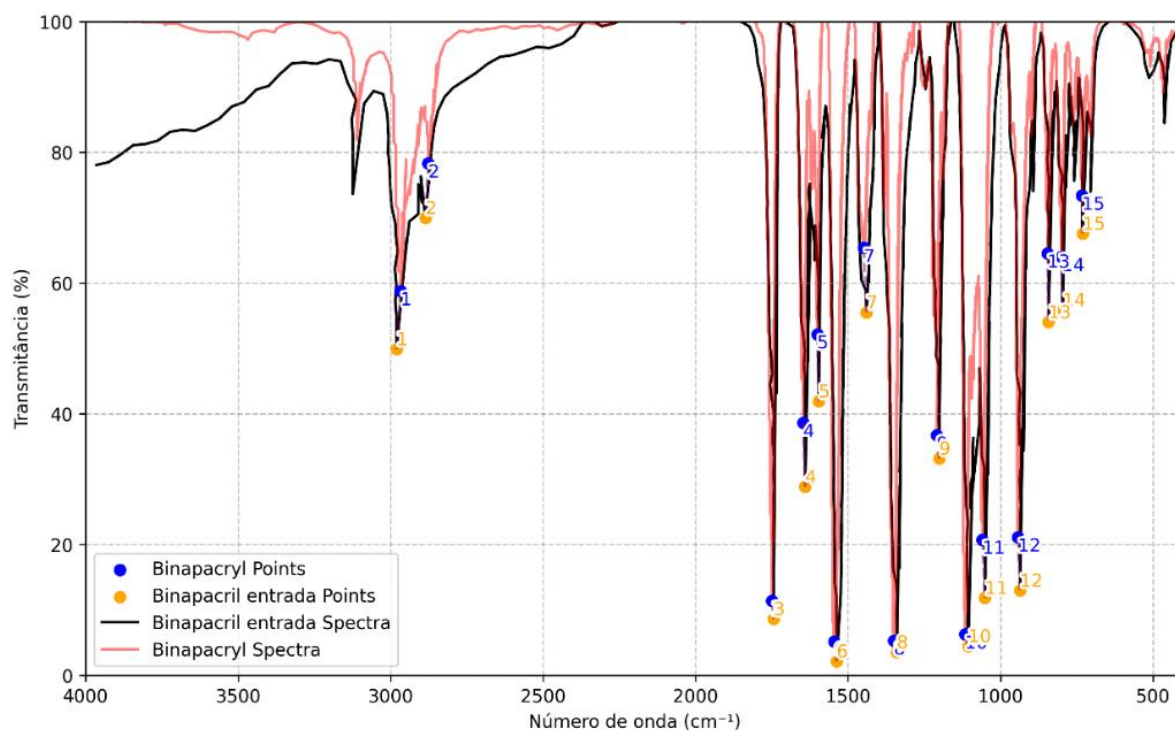
Figura 18 – Representação da estrutura química do Binapacril.



Fonte: Adaptado de PubChem, 2024d.

Discutido em seções anteriores, o relatório exibe uma sobreposição dos espectros IV, oferecendo uma perspectiva para o usuário sobre a eficácia do algoritmo em identificar bandas de interesse. A Figura 19 ilustra essa sobreposição para o Binapacril.

Figura 19 – Espectros IV sobrepostos do Binapacril do banco de dados (vermelho) e de entrada (preto).



Fonte: Autor.

Conforme analisado em tópicos anteriores, a Tabela 6 contém os valores de número de onda e transmitância para os pontos azul e laranja da Figura 19.

Tabela 6 – Comparação das bandas de interesse entre o Binapacril do banco de dados (Banco de dados) e o de entrada (Amostra).

Índex	Banco de dados	Amostra	(continua)	
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Transmitância (%)	Transmitância (%)
1	2967,96	2980,53	58,77	49,92
2	2876,62	2885,79	78,30	69,98
3	1750,13	1744,02	11,40	8,65
4	1647,10	1640,68	38,63	28,85
5	1601,14	1595,87	52,17	41,92
6	1545,24	1537,34	5,18	2,18
7	1448,93	1439,72	65,39	55,53
8	1349,36	1340,16	5,35	3,56
9	1209,58	1201,12	36,73	33,20

Índex	(conclusão)			
	Banco de dados	Amostra	Banco de dados	Amostra
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Transmitância (%)	Transmitância (%)
10	1116,84	1105,54	6,29	4,49
11	1059,80	1050,97	20,77	11,87
12	942,61	936,23	21,12	12,99
13	846,50	842,76	64,55	54,05
14	798,90	795,83	64,01	55,89
15	732,78	730,93	73,38	67,59

Fonte: Autor.

O algoritmo desenvolvido destacou o Binapacril como o composto de maior similaridade, alcançando um fator de similaridade de 91,30%. Seguiram-se Dinocape (61,46%), Fenitrotiona (60,72%), Fosadona (57,05%) e Fenoxaprope-p-etílico (56,00%). Todos os métodos de aprendizado de máquina utilizados — KNN, Regressão Logística e Naive-Bayes — também classificaram o Binapacril como o composto de maior similaridade. A Figura 20 destaca visualmente os resultados da classificação do Binapacril por cada um desses métodos. O Apêndice E contém as estruturas moleculares de todos os pesticidas.

Figura 20 – Classificação do Binapacril por ordem de similaridade (maior para menor).

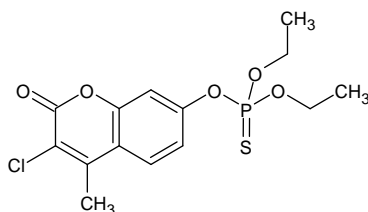
Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
Binapacril	Binapacril	Binapacril	Binapacril
Dinocape	Fluometurom	1,3-Dicloropropeno	Trifluralina
Fenitrotiona	1,3-Dicloropropeno	Linurom	Tetraclorvinfos
Fosadona	Linurom	Fluometurom	Fluometurom
Fenoxaprope-p-etílico	Cloroprofam	Lindano	1,3-Dicloropropeno

Fonte: Autor.

5.5.5 Coumafós

O Coumafós é empregado como inseticida e pesticida, conforme detalhado em PubChem (2024e). A estrutura química desse composto está apresentada na Figura 21.

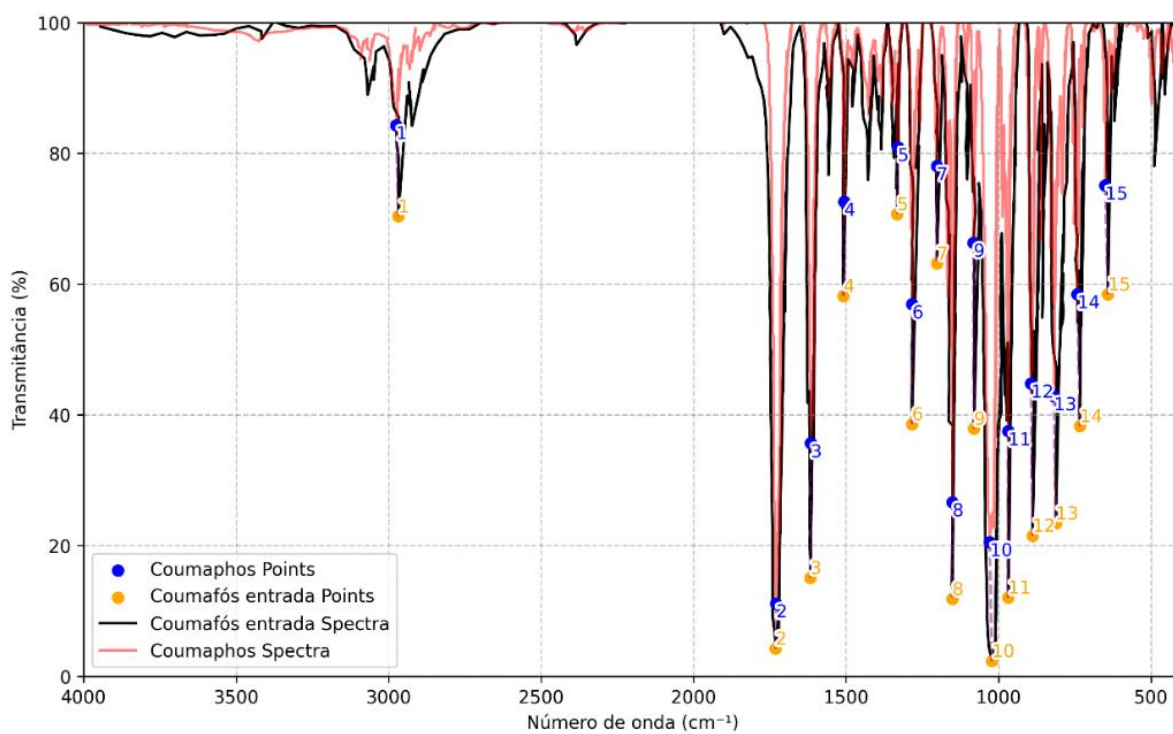
Figura 21 – Representação da estrutura química do Coumafós.



Fonte: Adaptado de PubChem, 2024e.

Como abordado em outras seções, a visualização no relatório apresenta uma comparação entre o espectro submetido e o do banco de dados. A Figura 22 mostra a sobreposição resultante dessa comparação para o Coumafós.

Figura 22 – Espectros IV sobrepostos do Coumafós do banco de dados (vermelho) e de entrada (preto).



Fonte: Autor.

Como já visto, a Tabela 7 exibe os detalhes dos valores de número de onda e transmitância para cada ponto azul e laranja da Figura 22.

Tabela 7 – Comparação das bandas de interesse entre o Coumafós do banco de dados (Banco de dados) e o de entrada (Amostra).

Índex	Banco de dados Número de onda (cm^{-1})	Amostra Número de onda (cm^{-1})	Banco de dados Transmitância (%)	Amostra Transmitância (%)
1	2975,37	2968,54	84,31	70,36
2	1729,30	1731,18	11,18	4,27
3	1615,98	1617,20	35,67	15,07
4	1506,36	1508,62	72,61	58,20
5	1330,79	1332,72	80,97	70,69
6	1283,15	1283,98	56,93	38,60
7	1202,73	1202,58	78,05	63,15
8	1151,34	1151,48	26,66	11,86
9	1081,94	1080,65	66,35	37,89
10	1028,76	1021,68	20,56	2,35
11	968,46	968,22	37,50	12,09
12	893,50	888,53	44,79	21,54
13	818,50	812,29	42,87	23,42
14	741,73	733,34	58,46	38,29
15	650,33	641,35	75,09	58,45

Fonte: Autor.

O algoritmo desenvolvido classificou o Coumafós como o composto de maior similaridade, atingindo um fator de similaridade de 98,20%. Na sequência, vieram Oxamil (56,94%), Monocrotofós (55,48%), Metidation (54,49%) e Menvifós (50,80%). Todos os métodos de aprendizado de máquina utilizados corroboram a classificação do Coumafós como o composto de maior similaridade. Os resultados detalhados de cada abordagem são exibidos na Figura 23, que proporciona uma representação visual clara das comparações entre os métodos. As geometrias dos agrotóxicos estão listadas no Apêndice F.

Figura 23 – Classificação do Coumafós por ordem de similaridade (maior para menor).

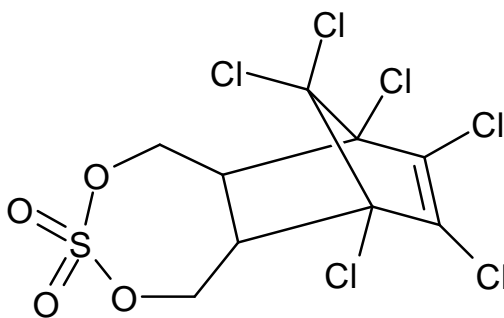
Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
Coumafós	Coumafós	Coumafós	Coumafós
Oxamil	Demeton	Hexaconazol	Demeton
Monocrotofós	Bromopropilato	Bromopropilato	Dissulfotom
Metidation	Captafol	Fosmete	Hexaconazol
Mevinfós	Dissulfotom	Demeton	Bromopropilato

Fonte: Autor.

5.5.6 Sulfato de endossulfam

De acordo com PubChem (2024f), o Sulfato de endossulfam é comumente utilizado como inseticida. Sua estrutura química está representada na Figura 24.

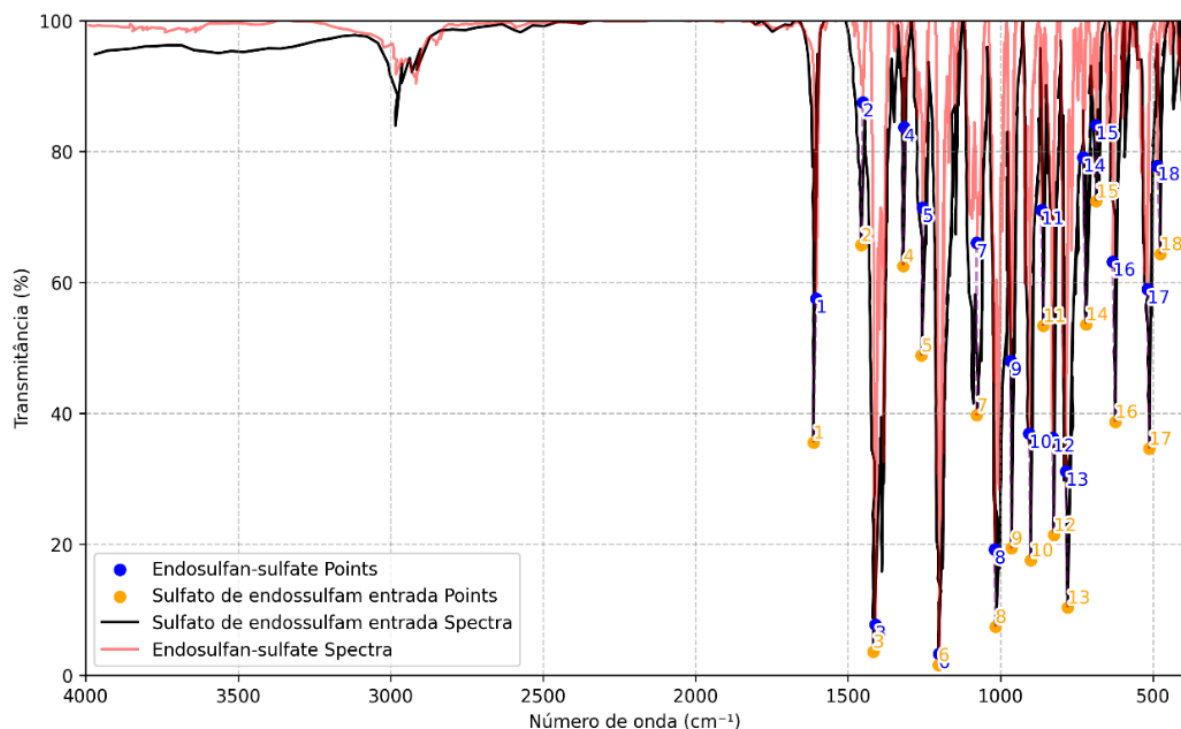
Figura 24 – Representação da estrutura química do Sulfato de endossulfam.



Fonte: Adaptado de PubChem, 2024f.

As seções prévias discutem como o relatório demonstra uma sobreposição dos espectros, proporcionando uma visualização de comparação entre os espectros do banco de dados e de entrada do usuário. A Figura 25 mostra essa visualização para o Sulfato de endossulfam.

Figura 25 – Espectros IV sobrepostos do sulfato de endossulfam do banco de dados (vermelho) e de entrada (preto).



Fonte: Autor.

Como estabelecido na Seção 5.5.1, a Tabela 8 fornece detalhes dos números de onda e transmitância correspondentes a cada ponto azul e laranja mostrado na Figura 25.

Tabela 8 – Comparação das bandas de interesse entre o sulfato de endossulfam do banco de dados (Banco de dados) e o de entrada (Amostra).

(continua)

Índex	Banco de dados	Amostra	Banco de dados	Amostra
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Transmitância (%)	Transmitância (%)
1	1605,08	1613,69	57,53	35,57
2	1451,53	1457,13	87,49	65,72
3	1411,02	1417,53	7,73	3,64
4	1316,16	1318,99	83,68	62,52
5	1255,75	1259,12	71,43	48,86
6	1202,48	1203,87	3,29	1,60
7	1078,31	1077,69	66,04	39,74
8	1019,62	1016,91	19,24	7,43

Índex	(conclusão)			
	Banco de dados	Amostra	Banco de dados	Amostra
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Transmitância (%)	Transmitância (%)
9	966,67	964,42	48,00	19,39
10	906,27	900,87	36,91	17,58
11	866,14	860,35	71,06	53,41
12	829,44	825,35	36,31	21,46
13	785,53	780,22	31,14	10,40
14	728,99	719,99	79,08	53,62
15	686,94	686,29	84,05	72,39
16	632,00	623,97	63,20	38,74
17	518,58	512,22	58,96	34,63
18	485,72	476,31	77,82	64,36

Fonte: Autor.

O algoritmo desenvolvido destacou o Sulfato de endossulfam como o composto de maior fator de similaridade, marcando 90,92%. Seguido por Diazinon (52,31%), Tricolorofom (51,04%), Fenitrothion (50,94%) e Hidrazida Maleica (50,14%). Com exceção do KNN, que classificou o Sulfato de endossulfam como o quarto mais similar, todos os outros métodos de aprendizado de máquina — Regressão Logística e Naive-Bayes — reconheceram-no como o composto de maior similaridade. A Figura 26 ilustra de maneira comparativa os resultados de classificação do Sulfato de endossulfam pelos diferentes métodos analíticos. No Apêndice G, estão as estruturas moleculares de todos os agrotóxicos.

Figura 26 – Classificação do sulfato de endossulfam por ordem de similaridade (maior para menor).

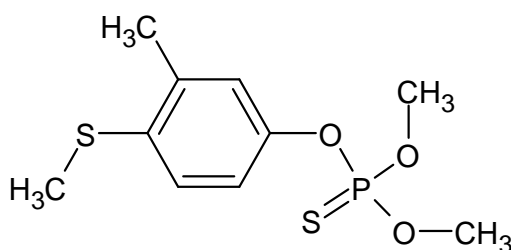
Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
Sulfato de endossulfam	Demeton	Sulfato de endossulfam	Sulfato de endossulfam
Diazinon	Dissulfotom	Dissulfotom	Demeton
Triclorofom	Difenoquat-metil-sulfato	Demeton	Dissulfotom
Fenitroion	Sulfato de endossulfam	Metidation	Etridiazol
Hidrazida maleica	Enridiazol	Clorpirifós-metil	Difenoquat-metil-sulfato

Fonte: Autor.

5.5.7 Fention

O Fention é utilizado tanto como inseticida quanto como pesticida, de acordo com informações do PubChem (2024g). A estrutura química deste composto é ilustrada na Figura 27.

Figura 27 – Representação da estrutura química do Fention.

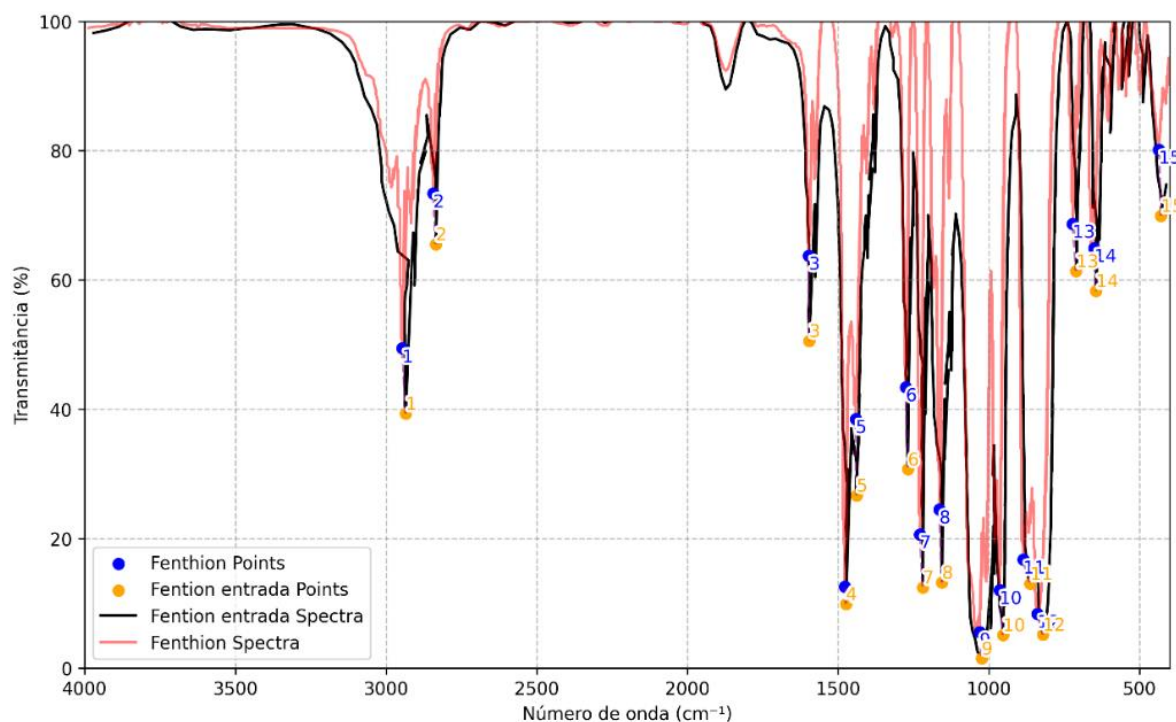


Fonte: Adaptado de PubChem, 2024g.

Como descrito nas seções anteriores, o relatório enfatiza uma sobreposição dos espectros do composto de referência com o analisado, facilitando uma análise detalhada das bandas escolhidas pelo algoritmo. A Figura 28 ilustra essa comparação para o Fention. As geometrias moleculares dos pesticidas estão no Apêndice H.

□

Figura 28 – Espectros IV sobrepostos do Fenthion do banco de dados (vermelho) e de entrada (preto).



Fonte: Autor.

Como estabelecido anteriormente na Seção 5.5.1, a Tabela 9 fornece detalhes dos números de onda e transmitância correspondentes a cada ponto azul e laranja mostrado na Figura 28.

Tabela 9 – Comparação das bandas de interesse entre o Fenthion do banco de dados (Banco de dados) e o de entrada (Amostra).

Índex	Banco de dados	Amostra	(continua)	
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Banco de dados Transmitância (%)	Amostra Transmitância (%)
1	2946,04	2934,72	49,44	39,35
2	2843,73	2835,28	73,40	65,54
3	1597,12	1596,22	63,77	50,60
4	1477,35	1474,23	12,62	9,94
5	1442,06	1438,79	38,52	26,73
6	1275,39	1268,80	43,45	30,74
7	1228,27	1219,22	20,70	12,50
8	1164,15	1155,20	24,55	13,24

Índex	Banco de dados Número de onda (cm ⁻¹)	Amostra Número de onda (cm ⁻¹)	(conclusão)	
			Banco de dados Transmitância (%)	Amostra Transmitância (%)
9	1030,83	1024,04	5,62	1,51
10	964,81	953,53	12,09	5,14
11	885,98	862,79	16,79	13,07
12	838,54	820,69	8,43	5,20
13	721,92	710,31	68,71	61,40
14	650,58	644,26	64,97	58,32
15	436,21	428,72	80,18	69,96

Fonte: Autor.

O algoritmo desenvolvido identificou o Fention como o composto mais similar, com um fator de similaridade de 99,52%. Este foi seguido por Thiram (65,13%), Fenitroton (64,08%), Clorpirifós-metil (62,96%) e Tetraclorvinfos (62,88%). Nenhum dos métodos de aprendizado de máquina foi capaz de identificar similaridade com o Fention. Os resultados de classificação do Fention de cada método são visualmente representados na Figura 29.

Figura 29 – Classificação do Fention por ordem de similaridade (maior para menor).

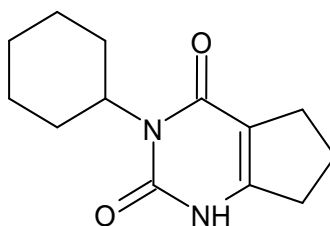
Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
Fention	Demeton	Metidation	Metidation
Thiram	Dissulfotom	Procimidona	Procimidona
Fenitroton	Difenzoquat-metil-sulfato	Tolifluanida	Demeton
Clorpirifós-metil	Metidation	1,3-Dicloropropeno(cis)	Dissulfotom
Tetraclorvinfos	Flusilazol	Hexaconazol	Difenzoquat-metil-sulfato

Fonte: Autor.

5.5.8 Lenacil

O Lenacil, utilizado como pesticida, possui uma ingestão diária aceitável (ADI) estabelecida em 0,12 mg por kg de peso corporal por dia, conforme as normas da União Europeia (PubChem, 2024h). Sua estrutura química pode ser visualizada na Figura 30.

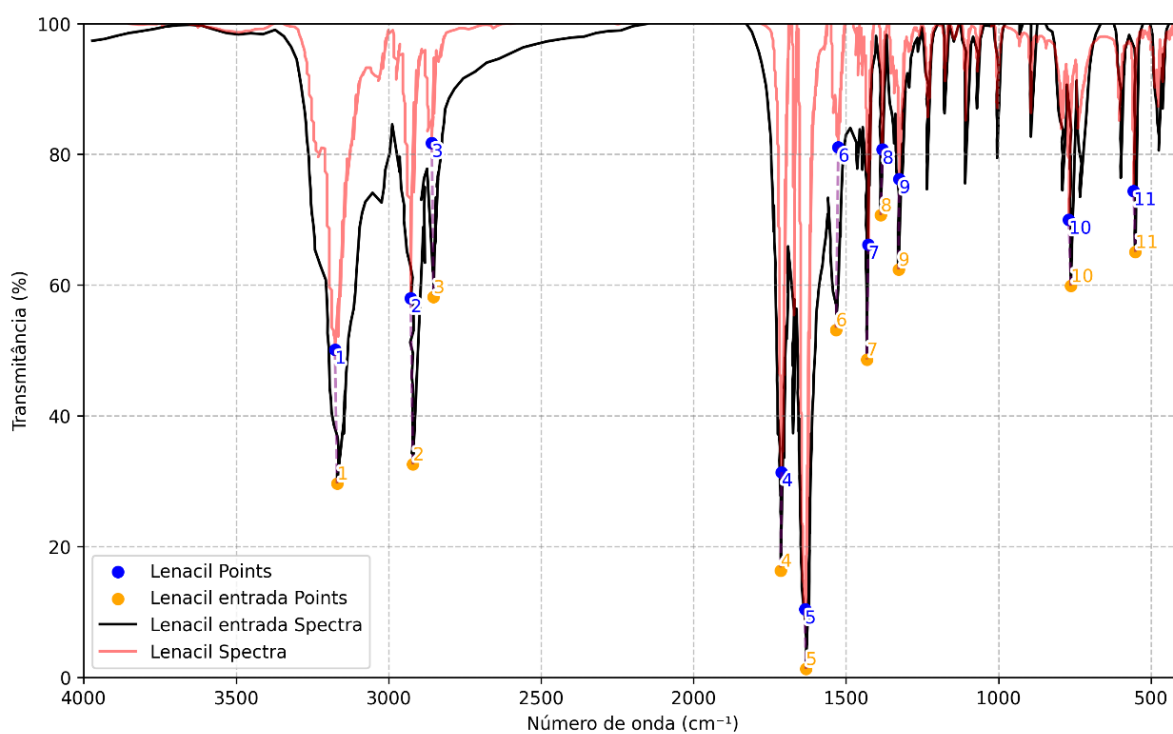
Figura 30 – Representação da estrutura química do Lenacil.



Fonte: Adaptado de PubChem, 2024h.

Em conformidade com o explorado nas seções prévias, a análise no relatório realça a sobreposição entre os espectros do composto de referência e o composto testado, permitindo um exame das bandas identificadas pelo algoritmo. A Figura 31 exibe essa sobreposição para o Lenacil.

Figura 31 – Espectros IV sobrepostos do Lenacil do banco de dados (vermelho) e de entrada (preto).



Fonte: Autor.

Seguindo a definição da Seção 5.5.1, a Tabela 10 lista os valores de número de onda e transmitância associados a cada ponto azul e laranja da Figura 31.

Tabela 10 – Comparação das bandas de interesse entre o Lenacil do banco de dados (Banco de dados) e o de entrada (Amostra).

Índex	Banco de dados Número de onda (cm^{-1})	Amostra Número de onda (cm^{-1})	Banco de dados Transmitância (%)	Amostra Transmitância (%)
1	3176,23	3168,97	50,13	29,67
2	2927,77	2920,21	58,02	32,62
3	2858,35	2853,02	81,77	58,18
4	1711,28	1714,63	31,40	16,35
5	1634,85	1631,97	10,43	1,31
6	1525,41	1533,20	81,06	53,15
7	1426,78	1432,13	66,19	48,65
8	1380,67	1386,43	80,73	70,69
9	1325,84	1328,03	76,24	62,39
10	769,45	762,83	70,02	59,89
11	557,02	551,23	74,38	65,09

Fonte: Autor.

O algoritmo desenvolvido classificou o Lenacil como o composto mais similar, alcançando um índice de similaridade de 83,28%. Os compostos que se seguiram foram Bromacil (61,09%), Dodina (49,95%), Cloroprofam (46,73%) e Metribuzim (46,38%). Com exceção do KNN, que não reconheceu o Lenacil em sua análise, todos os outros métodos de aprendizado de máquina — Regressão Logística e Naive-Bayes — identificaram o Lenacil como o composto de maior similaridade. A Figura 32 oferece uma representação visual detalhada, mostrando a comparação dos resultados de classificação do Lenacil pelos diversos métodos utilizados. Todas as geometrias dos agrotóxicos estão representadas no Apêndice I.

Figura 32 – Classificação do Lenacil por ordem de similaridade (maior para menor).

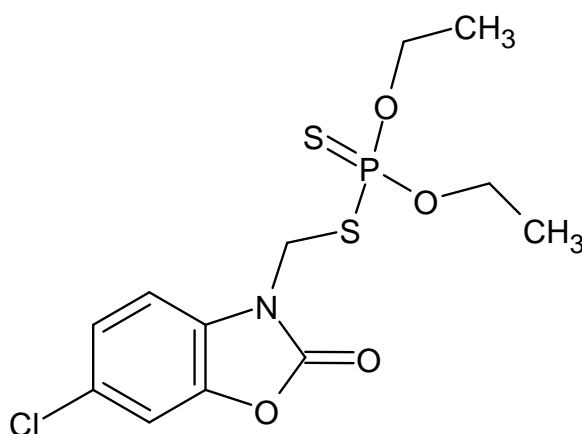
Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
Lenacil	Clorfimeforme	Lenacil	Lenacil
Bromacil	Bendiocarbe	Clorfimeforme	Clorfimeforme
Dodina	Carbofurano	Dodina	Penciurom
Cloroprofam	Dodina	Penciurom	Bendiocarbe
Metribuzim	Amitraz	Bendiocarbe	Carbofurano

Fonte: Autor.

5.5.9 Fosalona

A Fosalona é empregada como inseticida, pesticida e acaricida, conforme registrado em PubChem (2024i). A estrutura química deste composto está ilustrada na Figura 33.

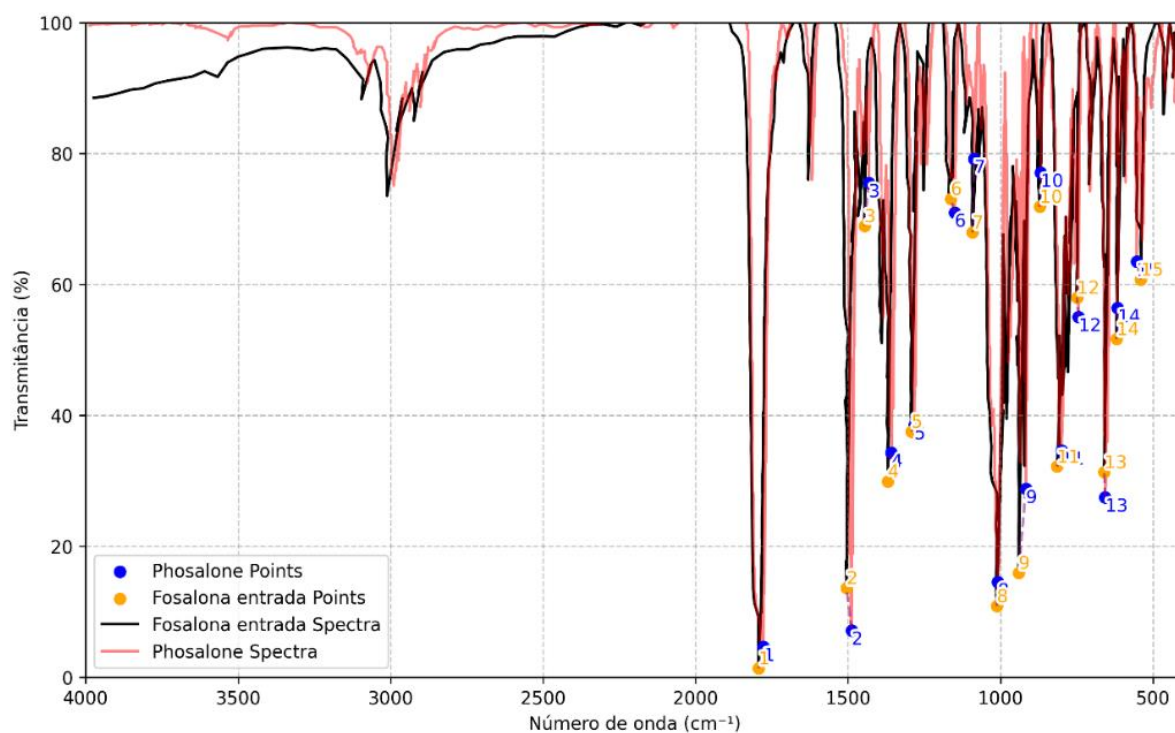
Figura 33 – Representação da estrutura química da Fosalona.



Fonte: Adaptado de PubChem, 2024i.

Como mencionado nas seções anteriores, o relatório apresenta uma sobreposição entre os espectros do composto referencial e o composto analisado, proporcionando ao usuário uma análise das bandas selecionadas pelo algoritmo. A Figura 34 evidencia essa comparação visual para a Fosalona.

Figura 34 – Espectros IV sobrepostos da Fosalone do banco de dados (vermelho) e de entrada (preto).



Fonte: Autor.

Como descrito previamente, os valores especificados de números de onda e transmitância representados por cada um dos pontos azul e laranja na Figura 34 são detalhados na Tabela 11.

Tabela 11 – Comparação das bandas de interesse entre a Fosalone do banco de dados (Banco de dados) e o de entrada (Amostra).

Índex	Banco de dados	Amostra	(continua)	
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Banco de dados Transmitância (%)	Amostra Transmitância (%)
1	1777,73	1793,22	4,67	1,39
2	1488,44	1504,64	7,17	13,65
3	1432,21	1446,28	75,51	68,92
4	1357,92	1369,54	34,28	29,94
5	1282,76	1292,34	38,53	37,54
6	1150,32	1163,36	70,99	73,09
7	1086,28	1092,38	79,20	67,97

Índex	Banco de dados	Amostra	Banco de dados	Amostra
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Transmitância (%)	Transmitância (%)
8	1008,73	1012,03	14,56	10,88
9	916,91	940,80	28,78	15,96
10	868,33	871,41	77,08	71,88
11	799,58	814,43	34,67	32,19
12	744,24	749,55	55,03	57,98
13	656,92	659,57	27,49	31,36
14	616,03	619,59	56,39	51,67
15	553,60	540,00	63,48	60,76

Fonte: Autor.

O algoritmo desenvolvido destacou a Fosalone como o composto mais similar, com um fator de similaridade de 98,58%. Seguiram-se Malation (54,88%), Triadimefom (54,40%), Dicamba (50,30%) e Fenitrothion (50,19%). Embora todos os métodos de aprendizado de máquina tenham reconhecido a Fosalone entre os compostos similares, houve variações nas suas classificações: KNN a colocou como terceira mais similar, Regressão Logística como segunda e Naive-Bayes como quinta. A Figura 35 ilustra essas diferenças nos resultados da classificação da Fosalone para cada um dos métodos analisados. Todas as geometrias dos pesticidas são apresentadas no Apêndice J.

Figura 35 – Classificação da Fosalone por ordem de similaridade (maior para menor).

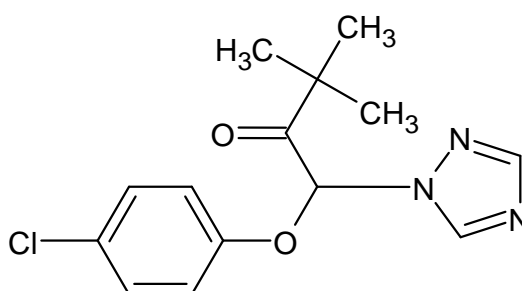
Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
Fosalone	Amitraz	Metidation	Amitraz
Malation	Difenoquat-metil-sulfato	Fosalone	Procimidona
Triadimefom	Fosalone	Propazina	Difenoquat-metil-sulfato
Dicamba	Metidation	Ciromazina	Propazina
Fenitroton	Alloxydium-sodium	Amitraz	Fosalone

Fonte: Autor.

5.5.10 Triadimefom

O Triadimefom, que serve tanto como fungicida quanto pesticida, tem uma ingestão diária aceitável (ADI) definida pelas diretrizes da União Europeia em 0,03 mg por kg de peso corporal diariamente (PubChem, 2024j). A estrutura química deste composto está representada na Figura 36.

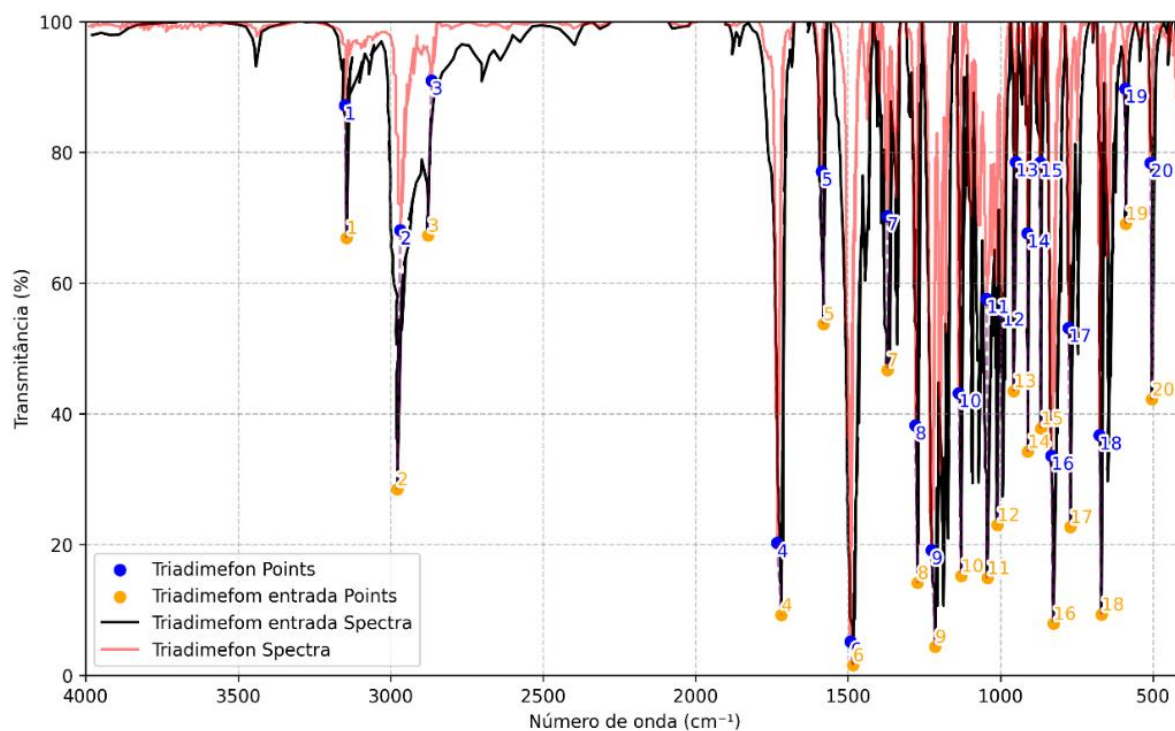
Figura 36 – Representação da estrutura química do Triadimefom.



Fonte: Adaptado de PubChem, 2024j.

Como abordado anteriormente, o relatório destaca uma sobreposição entre os espectros do composto de referência e o analisado, possibilitando uma análise por parte do usuário das bandas escolhidas pelo algoritmo. A Figura 37 demonstra essa comparação para o Triadimefom.

Figura 37 – Espectros IV sobrepostos do Triadimefom do banco de dados (vermelho) e de entrada (preto).



Fonte: Autor.

Conforme visto em seções anteriores, a Tabela 12 fornece uma especificação dos números de onda e transmitância para os pontos azul e laranja da Figura 37.

Tabela 12 – Comparação das bandas de interesse entre o Triadimefom do banco de dados (Banco de dados) e o de entrada (Amostra).

(continua)				
Índex	Banco de dados	Amostra	Banco de dados	Amostra
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Transmitância (%)	Transmitância (%)
1	3150,95	3144,45	87,16	66,93
2	2968,19	2978,65	68,10	28,47
3	2865,58	2877,83	90,98	67,28
4	1732,99	1718,79	20,25	9,25
5	1586,85	1580,62	77,08	53,75
6	1491,49	1484,22	5,19	1,58
7	1370,99	1370,60	70,28	46,71
8	1279,42	1272,39	38,27	14,16
9	1226,31	1214,43	19,18	4,40

Índex	(conclusão)			
	Banco de dados	Amostra	Banco de dados	Amostra
	Número de onda (cm ⁻¹)	Número de onda (cm ⁻¹)	Transmitância (%)	Transmitância (%)
10	1136,76	1129,32	43,19	15,23
11	1045,35	1043,51	57,59	14,89
12	994,13	1010,93	55,65	23,04
13	950,31	957,10	78,51	43,50
14	911,86	910,51	67,61	34,24
15	869,83	867,67	78,43	37,78
16	833,10	826,83	33,62	7,94
17	776,46	771,27	53,17	22,70
18	675,80	668,84	36,72	9,36
19	591,85	589,85	89,77	69,11
20	507,67	503,49	78,40	42,22

Fonte: Autor.

O algoritmo desenvolvido identificou o Triadimefom como o composto mais similar, com um fator de similaridade de 82,61%. Este foi seguido por Captafol (65,13%), Vinclozolina (59,59%), Pricimidona (58,43%) e Bendiocarbe (57,24%). Nenhum dos métodos de aprendizado de máquina foi capaz de identificar similaridade com o Triadimefom. Os resultados de classificação do Triadimefom de cada método são visualmente representados na Figura 38. O Apêndice K disponibiliza as geometrias moleculares de todos os agrotóxicos.

Figura 38 – Classificação do Triadimefom por ordem de similaridade (maior para menor).

Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
Triadimefom	Bitertanol	Isofenfos	Isofenfos
Captafol	Isofenfos	Diflubenzurom	Bitertanol
Vinclozolina	Demeton	Miclobutanil	Miclobutanil
Pricimidona	Dissulfotom	Triadimenol	Triadimenol
Bendiocarbe	Clorobenzilato	Bitertanol	Demeton

Fonte: Autor.

6 CONCLUSÕES E CONSIDERAÇÕES FINAIS

O desenvolvimento e a implementação de uma ferramenta interativa para identificação de espectros no infravermelho, foram elaborados a partir de uma interface em React.js com auxílio de CSS, suportada por uma API em Python, e integrando um banco de dados SQL para armazenamento de dados espectrais. O algoritmo de classificação desenvolvido foi otimizado para comparar espectros fornecidos pelos usuários com uma base de dados gratuita de espectros de agrotóxicos, fornecendo relatórios detalhados com comparações gráficas e tabulares dos cinco principais candidatos identificados.

Levando em conta somente a identificação do composto correto, o algoritmo desenvolvido classificou sem erros 100% dos compostos avaliados no estudo. Em comparação, os algoritmos de aprendizado de máquina apresentaram os seguintes desempenhos: KNN classificou corretamente 60% dos compostos, Regressão Linear 80%, e Naive-Bayes 70%. Considerando a identificação exata do composto de maior similaridade, o algoritmo desenvolvido também alcançou uma precisão de 100%. Os algoritmos de aprendizado de máquina utilizados no estudo apresentaram as seguintes precisões: KNN atingiu uma precisão de 40%, enquanto a Regressão Linear e o Naive-Bayes alcançaram uma precisão de 50% cada.

Embora o projeto tenha atingido suas metas fundamentais, ele apresenta algumas áreas que merecem atenção para refinamento. A principal é a falta de variedade de espectros provenientes de diferentes instrumentações, essa variável não foi totalmente explorada e poderia enriquecer os resultados da análise. Além disso, aprimoramentos na escalabilidade da ferramenta e na sua capacidade de processamento de dados em tempo real são necessários para garantir sua aplicabilidade em operações de maior escala.

Este trabalho estabeleceu uma integração de técnicas modernas de desenvolvimento, abrangendo todas as etapas que compõem uma aplicação funcional. Essa sinergia destacou a viabilidade da aplicação de abordagens computacionais em contextos científicos.

7 SUGESTÕES DE TRABALHOS FUTUROS

Para futuros desenvolvimentos, sugere-se a expansão do banco de dados para incluir uma faixa ainda mais ampla de espectros de substâncias químicas, introduzindo variações de espectros de um mesmo composto. Para que a expansão seja viável, seriam necessárias mais referências e realizar possíveis ajustes no procedimento de coleta e pré-tratamento de espectros, conforme a fonte. Essa etapa de expansão é possível de ser automatizada, mas apenas quando existir um padrão nos espectros de referência.

Uma abordagem alternativa, com foco na performance, seria criar bancos de dados específicos para grupos de compostos e adaptar a ferramenta para aceitar a troca flexível entre esses bancos de dados quando necessário. Essa estratégia otimiza o processamento, facilitando a busca e armazenamento de dados na memória de forma eficiente.

Além disso, aconselha-se aprimorar o algoritmo de classificação para aumentar sua precisão e eficiência para diferentes intensidades de bandas. Tal melhoria poderia incluir a reavaliação da linguagem de programação utilizada no desenvolvimento da API, buscando otimizar a performance e o tempo de execução da ferramenta.

Outra possibilidade seria adaptar o algoritmo para analisar diferentes tipos de espectros, como os de Raman e Ressonância Magnética Nuclear (RMN), que também se baseiam na identificação de picos e bandas. Esta adaptação poderia ser realizada com ajustes mínimos ao sistema já desenvolvido, necessitando apenas de um banco de dados apropriado e conjuntos de espectros específicos para teste.

8 REFERÊNCIAS

ANVISA. Agência Nacional de Vigilância Sanitária. **Publicada reclassificação toxicológica de agrotóxicos**. 2019. Disponível em: <<https://in.gov.br/web/dou/-/resolucao-re-n-2080-de-31-de-julho-de-2019-208203097>>. Acesso em 28 jun. 2024.

BAYU, A.; NANDIYANTO, D.; RAGADHITA, R.; FIANDINI, M. Interpretation of Fourier Transform Infrared Spectra (FTIR): A Practical Approach in the Polymer/Plastic Thermal Decomposition. **Indonesian Journal of Science and Technology**, v. 8, n. 1, p. 113-126, 2023.

BERKSON, J. Application of the logistic function to bio-assay. **Journal of the American statistical association**, v. 39, n. 227, p. 357-365, 1944.

BOHNER, T. O. L.; ARAÚJO, L. E. B.; NISHIJIMA, T. O impacto ambiental do uso de agrotóxicos no meio ambiente e na saúde dos trabalhadores rurais. **Revista eletrônica do curso de direito da UFSM**, v. 8, n. 3, p. 329-341, 2013.

BRASIL. **Lei nº 7802, de 11 de julho de 1989**. Disponível em: <https://www.planalto.gov.br/ccivil_03/leis/l7802.htm>. Acesso em 27 de jun. 2024.

BRASIL. **Lei nº 14785, de 27 de dezembro de 2023**. Disponível em: <https://www.planalto.gov.br/ccivil_03/_Ato2023-2026/2023/Lei/L14785.htm#art65>. Acesso em 27 jun. 2024.

CARVALHO, N. B. O.; LYRA, L. F.; SAKANE, K. K. Aplicação da espectroscopia vibracional no infravermelho, acoplada a análises estatísticas multivariadas, para a análise dos efeitos do uso de herbicidas à base do Diuron em feijões de tipo de grão carioca. **Research, Society and Development**, v. 12, n. 11, p. e136121143753-e136121143753, 2023.

COLTHUP, N.; DALY, L. H.; WIBERLEY S.E. **Introduction to infrared and Raman spectroscopy**. 3rd ed. Elsevier, 2012.

DA ROCHA, R. R.; DOS SANTOS, L. F. U.; PINTO, A. E. M. Retrocessos normativos no procedimento de registro de agrotóxicos no Brasil e o direito ao meio ambiente sadio. **Boletim do Observatório Ambiental Alberto Ribeiro Lamego**, v. 16, n. 2, p. 17-38, 2022.

DA SILVA, E. P. *et al.* Compatibility study between atorvastatin and excipients using DSC and FTIR. **Journal of Thermal Analysis and Calorimetry**, v. 123, p. 933-939, 2016.

DE ITOZ, M. A. V.; PEREIRA, K. C. AGROTÓXICO NO DIREITO AGRÁRIO E AMBIENTAL: UMA ANÁLISE DAS NORMAS REGULADORAS EM 2023. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 10, n. 5, p. 1517-1530, 2024.

DO NASCIMENTO, B. P. *et al.* Impacto ambiental sobre a saúde humana devido à exposição aos agrotóxicos. **UNICIÊNCIAS**, v. 25, n. 1, p. 44-56, 2021.

ENDERS, A. A. *et al.* Functional group identification for FTIR spectra using image-based machine learning models. **Analytical Chemistry**, v. 93, n. 28, p. 9711-9718, 2021.

GOLDBERGER, J. *et al.* Neighbourhood components analysis. **Advances in neural information processing systems**, v. 17, 2004.

GRIGORIU, N. *et al.* Monitoring the breakdown of Dinocap in spiked soil, wine and grape samples by GC/MS and FTIR/ATR. **Environmental Engineering & Management Journal (EEMJ)**, v. 17, n. 9, 2018.

IBAMA. Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis. **Portaria Normativa nº 84 de 15 de outubro de 1996**. Disponível em: <<https://www.ibama.gov.br/component/legislacao/?view=legislacao&legislacao=99498>>. Acesso em 28 jun. 2024.

LONDRES, F. Agrotóxicos no Brasil: um guia para ação em defesa da vida. Rio de Janeiro: **Rede Brasileira de Justiça Ambiental**, 2012.

MENDES, E.; DUARTE, N. Mid-Infrared Spectroscopy as a Valuable Tool to Tackle Food Analysis: A Literature Review on Coffee, Dairies, Honey, Olive Oil and Wine. **Foods**, v. 10, n. 2, p. 477, 2021.

National Center for Biotechnology Information. PubChem Compound Summary for CID 6862, 1-Naphthaleneacetic acid. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/compound/1-Naphthaleneacetic-acid>>. Acesso em: 30 jun. 2024a.

National Center for Biotechnology Information. PubChem Compound Summary for CID 91634, Azocyclotin. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/compound/Azocyclotin>>. Acesso em: 30 jun. 2024b.

National Center for Biotechnology Information. PubChem Compound Summary for CID 2314, Bendiocarb. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/compound/Bendiocarb>>. Acesso em: 30 jun. 2024c.

National Center for Biotechnology Information. PubChem Compound Summary for CID 10234, Binapacryl. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/compound/Binapacryl>>. Acesso em: 30 jun. 2024d.

National Center for Biotechnology Information. PubChem Compound Summary for CID 2871, Coumaphos. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/compound/Coumaphos>>. Acesso em: 30 jun. 2024e.

National Center for Biotechnology Information. PubChem Compound Summary for CID 13940, Endosulfan sulfate. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/compound/Endosulfan-sulfate>>. Acesso em: 30 jun. 2024f.

National Center for Biotechnology Information. PubChem Compound Summary for CID 3346, Fenthion. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/compound/Fenthion>>. Acesso em: 30 jun. 2024g.

National Center for Biotechnology Information. PubChem Compound Summary for CID 16559, Lenacil. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/compound/Lenacil>>. Acesso em: 30 jun. 2024h.

National Center for Biotechnology Information. PubChem Compound Summary for CID 4793, Phosalone. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/compound/Phosalone>>. Acesso em: 30 jun. 2024i.

National Center for Biotechnology Information. PubChem Compound Summary for CID 39385, Triadimefon. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/compound/Triadimefon>>. Acesso em: 30 jun. 2024j.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. Exposure to Highly Hazardous Pesticides: a Major Public Health Concern. Genebra: **WHO**, 2010.

PEREIRA, D. C.; PUPIN, B.; SAKANE, K. K. Evaluation of hydrogel use in the development of *Rapanea ferruginea* with water restriction by vibrational Fourier Transform Infrared Spectroscopy (FTIR-UATR). **Revista Ambiente & Água**, v. 16, n. 5, e2744, 2021.

PERES, F.; MOREIRA, J.C.; DUBOIS, G.S. Agrotóxicos, saúde e ambiente: uma introdução ao tema. É veneno ou remédio? Rio de Janeiro: **Fiocruz**, 2003.

PUGLIESE, R.; REGONDI, S.; MARINI, R. Machine learning-based approach: Global trends, research directions, and regulatory standpoints. **Data Science and Management**, v. 4, p. 19-29, 2021.

RISH, I. An empirical study of the naive Bayes classifier. In: **IJCAI 2001 workshop on empirical methods in artificial intelligence**. p. 41-46, 2001.

SAMPAIO, C. Nova classificação de agrotóxicos é "forma de enganar a sociedade", diz pesquisador: Metodologia muda rótulo dos produtos; Greenpeace aponta que sistematização confunde consumidores. 2019. Brasil de Fato. Disponível em: <<https://www.brasildefato.com.br/2019/07/24/especialistas-criticam-nova-classificacao-de-agrotoxicos-da-anvisa/>>. Acesso em 28 jun. 2024.

SANTANA, V. S.; MOURA, M. C. P.; NOGUEIRA, F. F. e. **Mortalidade por intoxicação ocupacional relacionada a agrotóxicos, 2000-2009**, Brasil. Revista de Saúde Pública, v. 47, n. 03, p. 598-606, 2013.

SECRETARIA NACIONAL DE VIGILÂNCIA SANITÁRIA. Ministério da Saúde. **Portaria nº 03 de 16 de janeiro de 1992**. Disponível em: <https://bvsms.saude.gov.br/bvs/saudelegis/svs1/1992/prt0003_16_01_1992.html>. Acesso em 28 jun. 2024.

SILVERSTEIN, R. M. *et al.* **Spectrometric Identification of Organic Compounds**. 8th ed. EUA: John Wiley & Sons, 2015.

VEIGA, M. M. *et al.* Análise da contaminação dos sistemas hídricos por agrotóxicos numa pequena comunidade rural do Sudeste do Brasil. **Caderno de Saúde Pública**. vol.22 nº.11 Rio de Janeiro, p. 2391-2399, 2006.

VISSER, T. **Infrared spectra of pesticides**. CRC Press, 2020.

ZHANG, H. The Optimality of Naive Bayes. In: **FLAIRS**. California: AAAI Press. p. 562-567, 2004.

ZHANG, Z.-M.; CHEN, S.; LIANG, Y.-Z. Baseline correction using adaptive iteratively reweighted penalized least squares. **Analyst**, v. 135, n. 5, p. 1138-1146, 2010.

APÊNDICE A – COMPONENTES DO BANCO DE DADOS

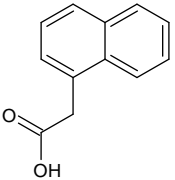
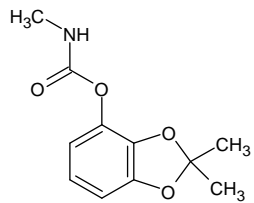
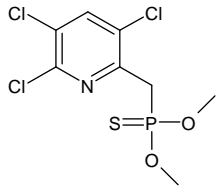
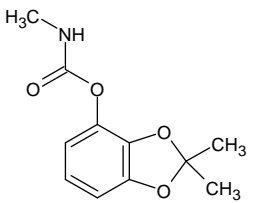
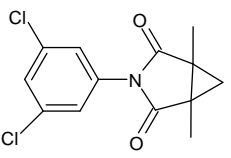
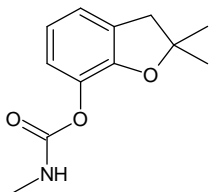
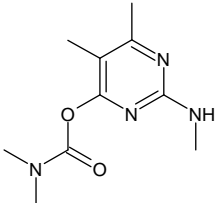
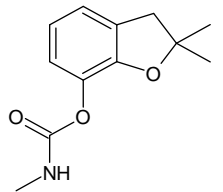
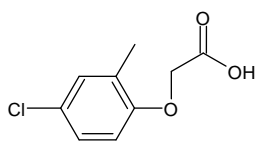
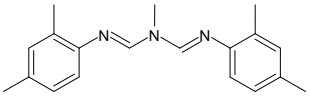
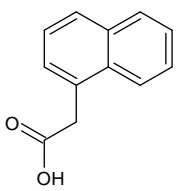
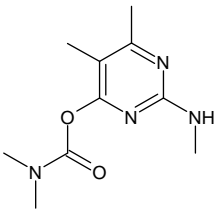
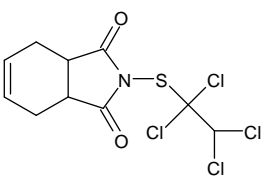
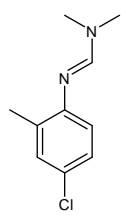
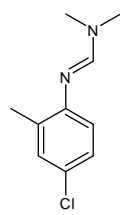
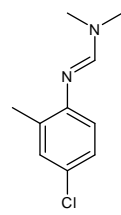
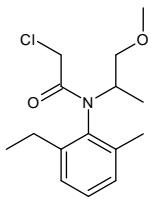
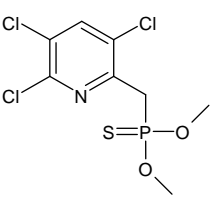
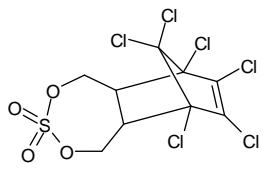
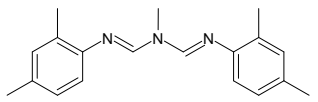
Id	Nome
1	1,2-Dichloropropane
2	1,3-Dichloropropene(cis)
3	1-Naphthol
4	1-Naphthyl-acetic-acid
5	2,4'-DDT
6	2,4,5-Trichlorophenol
7	2,4-DB
8	2,6-Dichlorobenzamide
9	4,4'-DDT
10	Acephate
11	Alachlor
12	Aldicarb
13	Aldrin
14	Alloxydim-sodium
15	Amitraz
16	Amitrole
17	Asulam
18	Atrazin
19	Azocyclotin
20	Bendiocarb
21	Bentazone
22	Benzthiazuron
23	Binapacryl
24	Bioresmethrin
25	Biphenyl
26	Bitertanol
27	Bromacil
28	Bromopropylate
29	Bronopol
30	Buprofezin
31	Captafol
32	Captan
33	Carbofuran
34	Carbosulfan
35	Chinomethionate
36	Chlordimeform
37	Chlorobenzilate
38	Chlorothiophos
39	Chlorpropham
40	Chlorpyrifos-methyl
41	Chlorpyrifos
42	Clofentezine
43	Coumaphos
44	Cyanazine
45	Cyanofenphos

Id	Nome
46	Cycloxydim
47	Cyhexatin
48	Cymoxanil
49	Cyproconazole
50	Cyromazine
51	Demeton
52	Desmedipham
53	Diazinon
54	Dicamba
55	Dichlofenthion
56	Dichlofluanid
57	Dichlorvos
58	Dicofol
59	Difenzoquat-methyl-sulphate
60	Diflubenzuron
61	Dimethipin
62	Dimethoate
63	Dinocap
64	Diphenylamine
65	Disulfoton
66	Diuron
67	Dodine
68	Endosulfan-sulfate
69	Endrin
70	Ethiofencarb
71	Ethoxyquin
72	Ethylene-thiourea
73	Etridiazole
74	Fenchlorpho
75	Fenitrothion
76	Fenoxaprop-ethyl
77	Fenoxycarb
78	Fenthion
79	Fenvalerate
80	Fluometuron
81	Flusilazole
82	Flutolanil
83	Folpeter
84	Formetanate
85	Formothion
86	Gibberellic-acid
87	Glufosinate-ammonium
88	Hexaconazole
89	Hexythiazox
90	Imazalil

Id	Nome
91	Imazapyr
92	Iprodione
93	Isofenphos
94	Lenacil
95	Lindane
96	Linuron
97	Malathion
98	Maleic-hydrazide
99	Maneb
100	MCPA
101	Metalaxyl
102	Methabenzthiazuron
103	Methamidophos
104	Methidathion
105	Methomy
106	Metolachlor
107	Metribuzin
108	Metsulfuron-methyl
109	Mevinphos
110	Monocrotophos
111	Monuron
112	Myclobutanil
113	Naled
114	Nicotine
115	Nitrofen
116	O-Phenylphenol
117	Oxamyl
118	Paraquat
119	Pencycuron
120	Pentachlorophenol
121	Phenmedipham
122	Phenthoate
123	Phosalone
124	Phosmet
125	Phoxim
126	Piperonyl-butoxide
127	Pirimicarb
128	PP-333
129	Prochloraz
130	Procymidone
131	Profenofos
132	Prometryne
133	Propazine
134	Propoxur
135	Propyzamide
136	Quinalphos

Id	Nome
137	Rotenone
138	Sethoxydim
139	Streptomycin
140	Tetrachlorvinphos
141	Tetradifon
142	Tetraphenyltin
143	Thiram
144	Tolylfluanid
145	Triadimefon
146	Triadimenol
147	Trichlorfon
148	Triclopyr
149	Trifluralin
150	Triforine
151	Vinclozolin

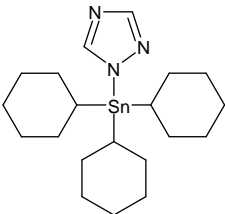
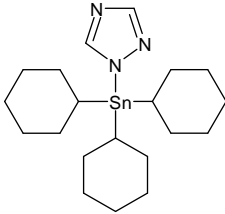
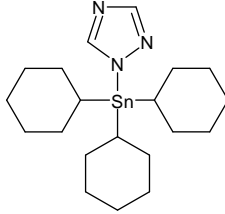
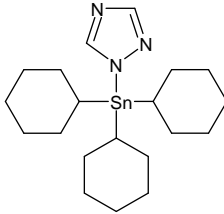
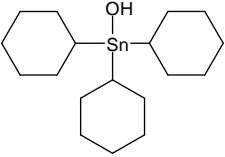
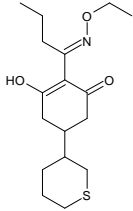
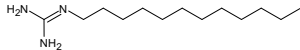
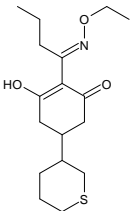
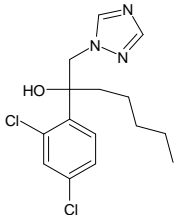
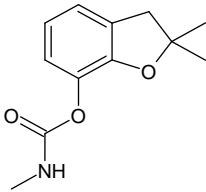
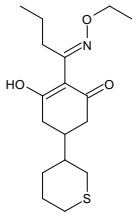
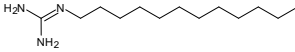
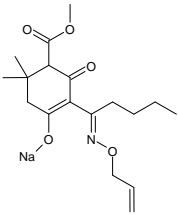
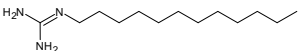
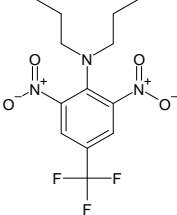
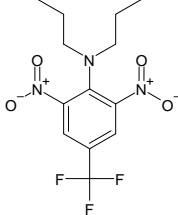
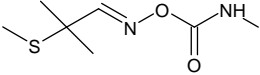
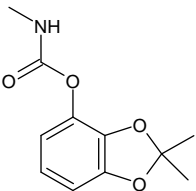
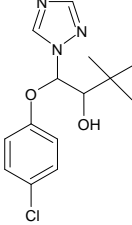
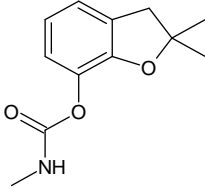
APÊNDICE B – ESTRUTURAS DOS RESULTADOS DO ÁCIDO 1-NAFTALENOACÉTICO

Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
1) <u>Ácido 1-naftalenoacético</u>	1) Bendiocarbe	1) Clorpirifós-metil	1) Bendiocarbe
			
2) Procimidona	2) Carbofurano	2) Pirimicarbe	2) Carbofurano
			
3) MCPA	3) Amitraz	3) <u>Ácido 1-naftalenoacético</u>	3) Pirimicarbe
			
4) Captafol	4) Clordimeforme	4) Clordimeforme	4) Clodimerforme
			
5) Metolaclo	5) Clorpirifós-metil	5) Sulfato de endossulfam	5) Amitraz
			

Fonte: Adaptado de NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. PubChem.

Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/>>. Acesso em: 30 jul. 2024.

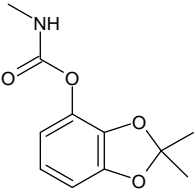
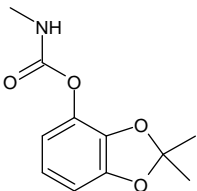
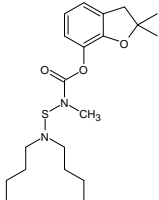
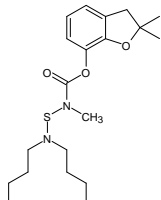
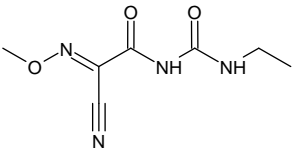
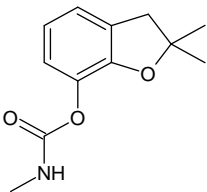
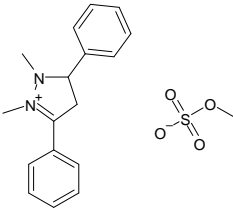
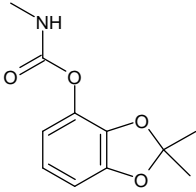
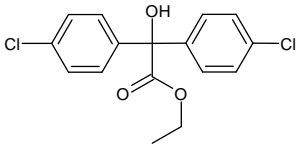
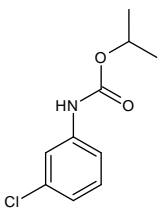
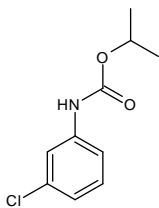
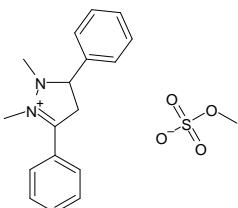
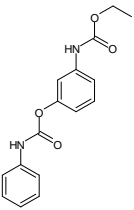
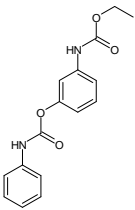
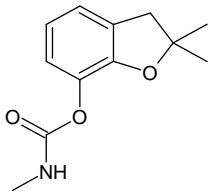
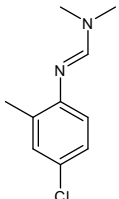
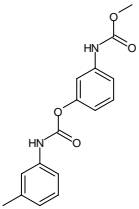
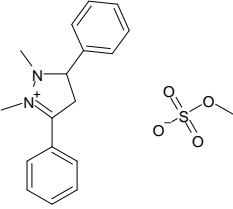
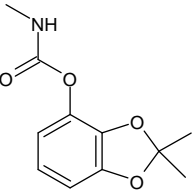
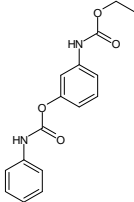
APÊNDICE C – ESTRUTURAS DOS RESULTADOS DA AZOCICLOTINA

Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
1) <u>Azociclotina</u> 	1) <u>Azociclotina</u> 	1) <u>Azociclotina</u> 	1) <u>Azociclotina</u> 
2) Cihexatina 	2) Cicloxidime 	2) Dodina 	2) Cicloxidime 
3) Hexaconazol 	3) Carbofurano 	3) Cicloxidime 	3) Dodina 
4) Alloxidum-sodium 	4) Dodina 	4) Trifluralina 	4) Trifluralina 
5) Aldicarbe 	5) Bendiocarbe 	5) Triadimenol 	5) Carbofurano 

Fonte: Adaptado de NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. PubChem.

Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/>>. Acesso em: 30 jul. 2024.

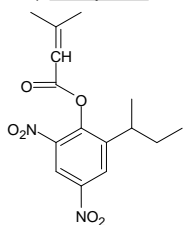
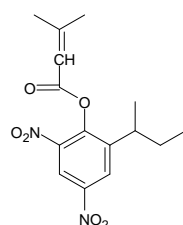
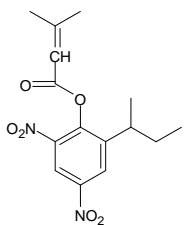
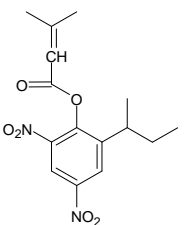
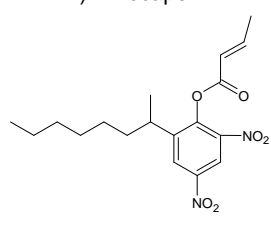
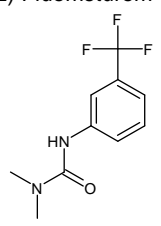
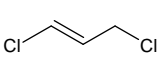
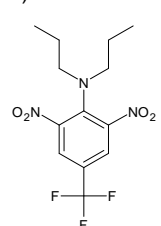
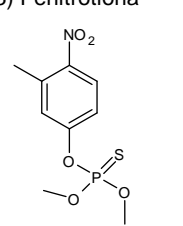
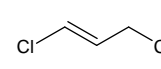
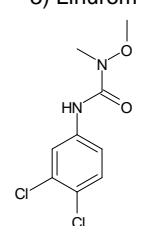
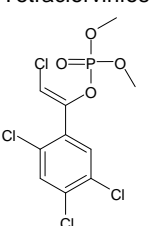
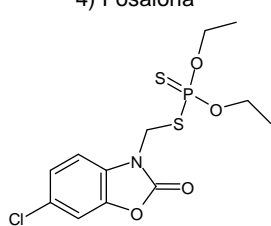
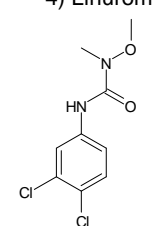
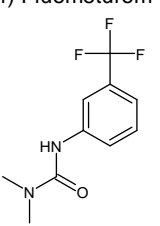
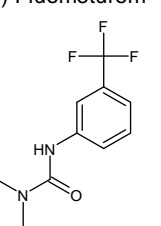
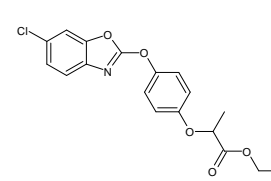
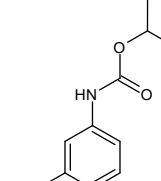
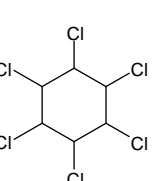
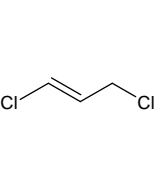
APÊNDICE D – ESTRUTURAS DOS RESULTADOS DO BENDIOCARBE

Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
1) <u>Bendiocarbe</u>	1) <u>Bendiocarbe</u>	1) Carbosulfano	1) Carbosulfano
			
2) Cimoxanil	2) Carbofurano	2) Difenzoquat-metil-sulfato	2) <u>Bendiocarbe</u>
			
3) Clorobezilato	3) Cloroprofam	3) Cloroprofam	3) Difenzoquat-metil-sulfato
			
4) Desmedifam	4) Desmedifam	4) Carbofurano	4) Cloroprofam
			
5) Fenmedifam	5) Difenzoquat-metil-sulfato	5) <u>Bendiocarbe</u>	5) Desmedifam
			

Fonte: Adaptado de NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. PubChem.

Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/>>. Acesso em: 30 jul. 2024.

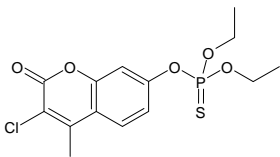
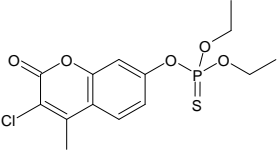
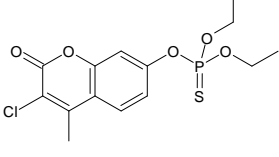
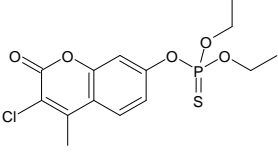
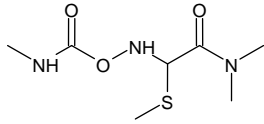
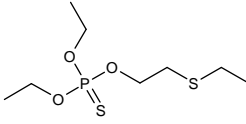
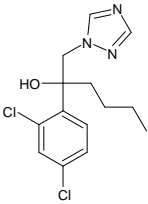
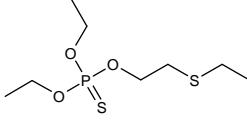
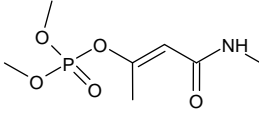
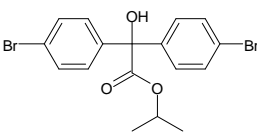
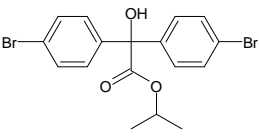
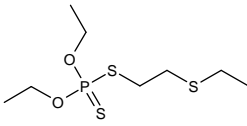
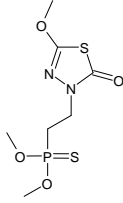
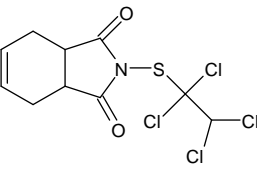
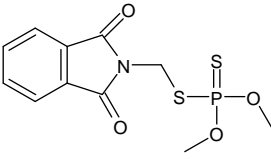
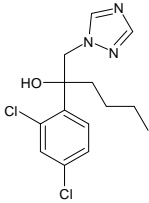
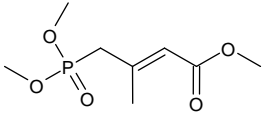
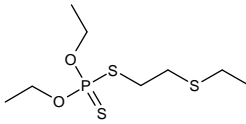
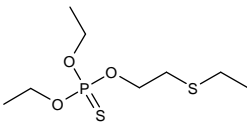
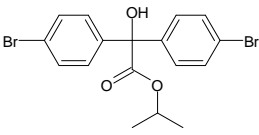
APÊNDICE E – ESTRUTURAS DOS RESULTADOS DO BINAPACRIL

Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
1) Binapacril 	1) Binapacril 	1) Binapacril 	1) Binapacril 
2) Dinocape 	2) Fluometurom 	2) 1,3-Dicloropropeno 	2) Trifluralina 
3) Fenitrotiona 	3) 1,3-Dicloropropeno 	3) Linurom 	3) Tetraclorvinfos 
4) Fosalona 	4) Linurom 	4) Fluometurom 	4) Fluometurom 
5) Fenoxaprop-p-etílico 	5) Cloroprofam 	5) Lindano 	5) 1,3-Dicloropropeno 

Fonte: Adaptado de NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. PubChem.

Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/>>. Acesso em: 30 jul. 2024.

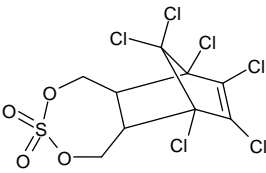
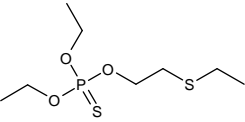
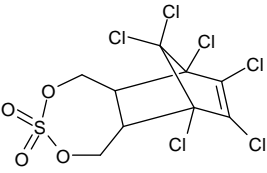
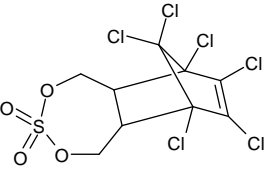
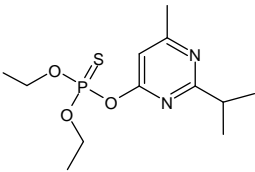
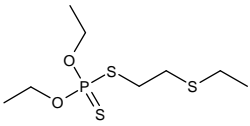
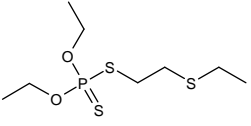
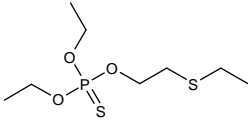
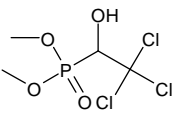
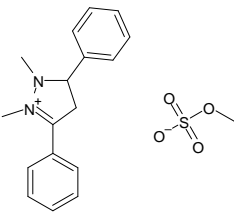
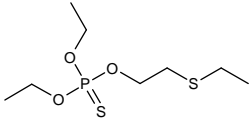
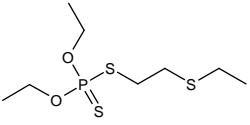
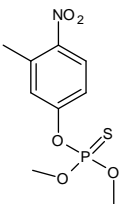
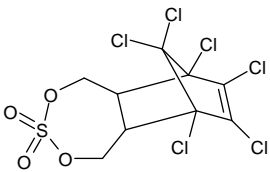
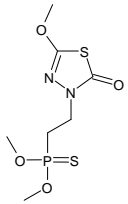
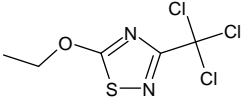
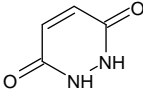
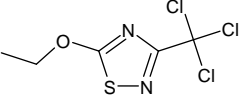
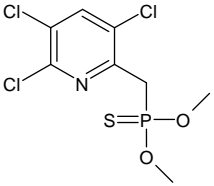
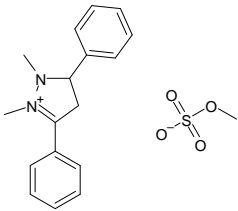
APÊNDICE F – ESTRUTURAS DOS RESULTADOS DO COUMAFÓS

Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
1) <u>Coumafós</u>	1) <u>Coumafós</u>	1) <u>Coumafós</u>	1) <u>Coumafós</u>
			
2) Oxamil	2) Demeton	2) Hexaconazol	2) Demeton
			
3) Monocrotofós	3) Bromopropilato	3) Bromopropilato	3) Dissulfotom
			
4) Metidation	4) Captafol	4) Fosmete	4) Hexaconazol
			
5) Mevinfós	5) Dissulfotom	5) Demeton	5) Bromopropilato
			

Fonte: Adaptado de NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. PubChem.

Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/>>. Acesso em: 30 jul. 2024.

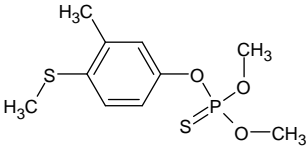
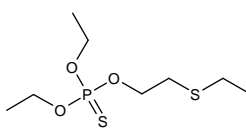
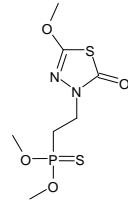
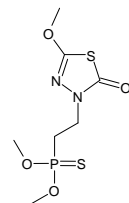
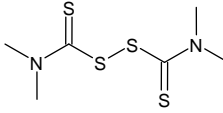
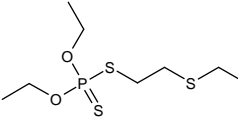
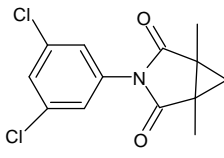
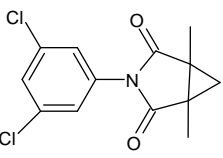
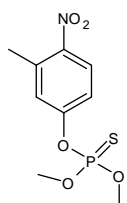
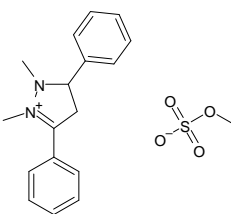
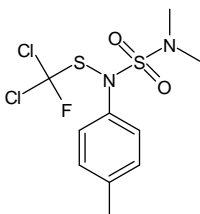
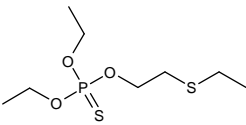
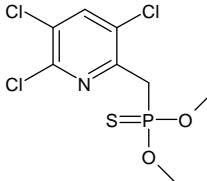
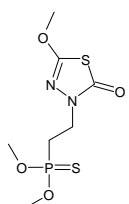
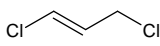
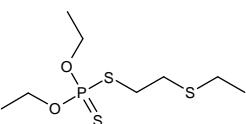
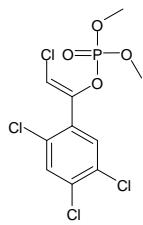
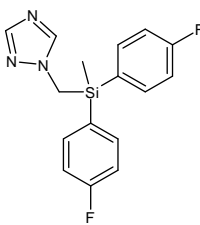
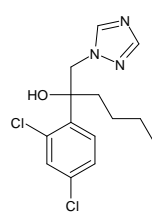
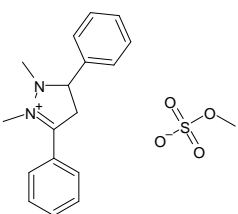
APÊNDICE G – ESTRUTURAS DOS RESULTADOS DO SULFATO DE ENDOSSULFAM

Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
1) <u>Sulfato de endossulfam</u>	1) Demeton	1) <u>Sulfato de endossulfam</u>	1) <u>Sulfato de endossulfam</u>
			
2) Diazinon	2) Dissulfotom	2) Dissulfotom	2) Demeton
			
3) Triclorofom	3) Difenzoquat-metil-sulfato	3) Demeton	3) Dissulfotom
			
4) Fenitrotona	4) <u>Sulfato de endossulfam</u>	4) Metidation	4) Etridiazol
			
5) Hidrazida maleica	5) Entridiazol	5) Clorpirifós-metil	5) Difenzoquat-metil-sulfato
			

Fonte: Adaptado de NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. PubChem.

Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/>>. Acesso em: 30 jul. 2024.

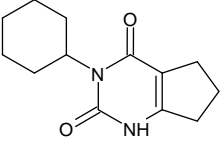
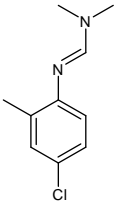
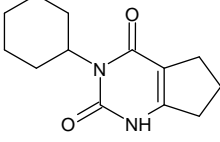
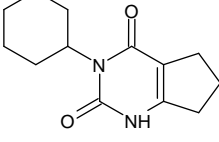
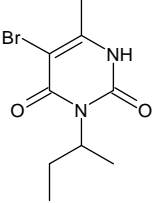
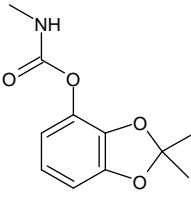
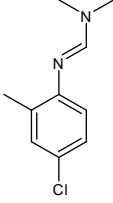
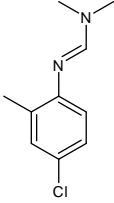
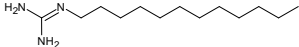
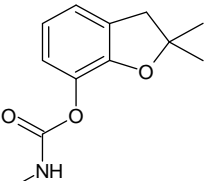
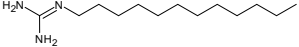
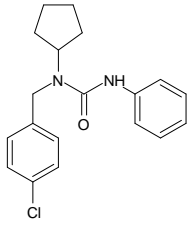
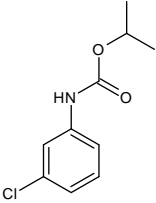
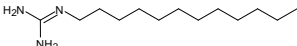
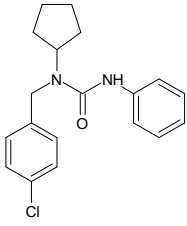
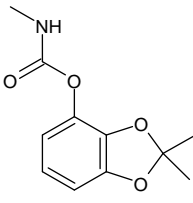
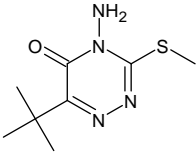
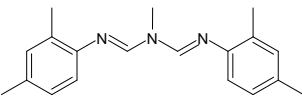
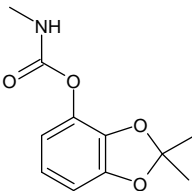
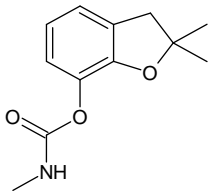
APÊNDICE H – ESTRUTURAS DOS RESULTADOS DO FENTION

Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
1) <u>Fention</u>	1) Demeton	1) Metidation	1) Metidation
			
2) Thiram	2) Dissulfotom	2) Procimidona	2) Procimidona
			
3) Fenitrotiona	3) Difenzoiquat-metil-sulfato	3) Tolifluanida	3) Demeton
			
4) Clorpirifós-metil	4) Metidation	4) 1,3-Dicloropropeno(cis)	4) Dissulfotom
			
5) Tetraclorvinfos	5) Flusilazol	5) Hexaconazol	5) Difenzoiquat-metil-sulfato
			

Fonte: Adaptado de NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. PubChem.

Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/>>. Acesso em: 30 jul. 2024.

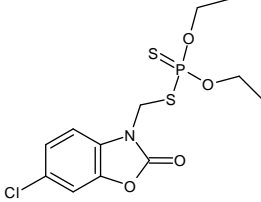
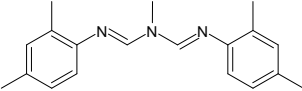
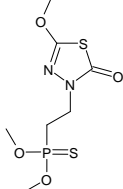
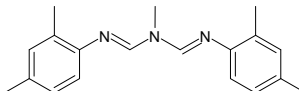
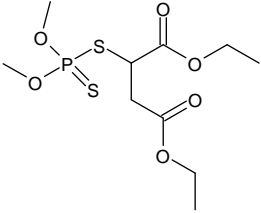
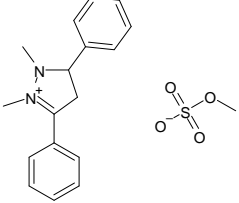
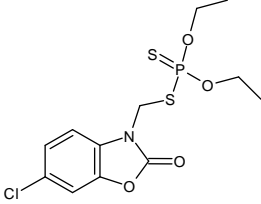
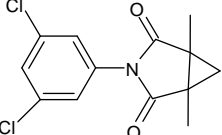
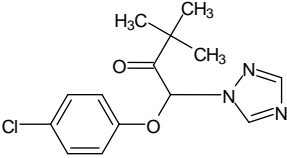
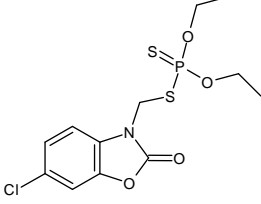
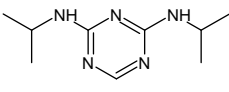
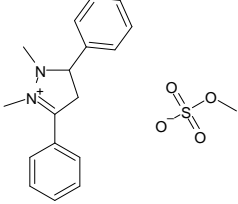
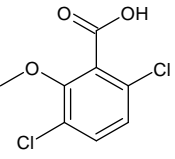
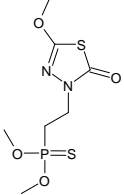
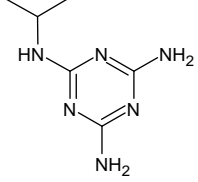
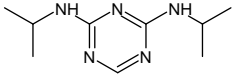
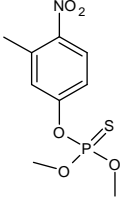
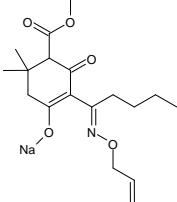
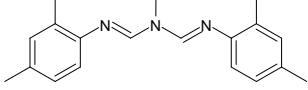
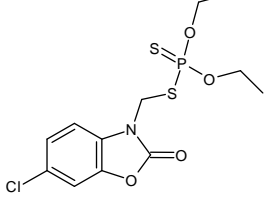
APÊNDICE I – ESTRUTURAS DOS RESULTADOS DO LENACIL

Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
1) <u>Lenacil</u>	1) Clorodimeforme	1) <u>Lenacil</u>	1) <u>Lenacil</u>
			
2) Bromacil	2) Bendiocarbe	2) Clordimeforme	2) Clordimeforme
			
3) Dodina	3) Carbofurano	3) Dodina	3) Pencicurom
			
4) Cloroprofam	4) Dodina	4) Pencicurom	4) Bendiocarbe
			
5) Metribuzim	5) Amitraz	5) Bendiocarbe	5) Carbofurano
			

Fonte: Adaptado de NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. PubChem.

Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/>>. Acesso em: 30 jul. 2024.

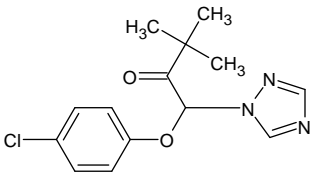
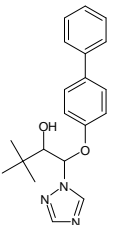
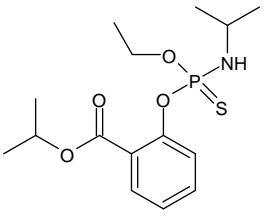
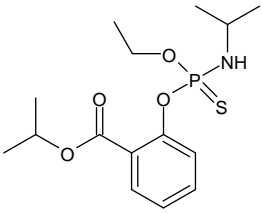
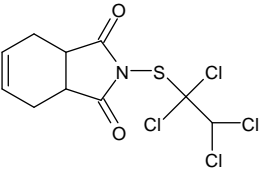
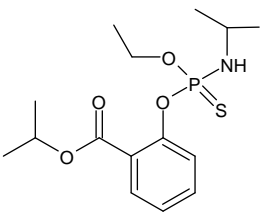
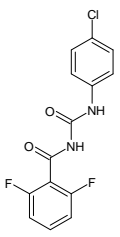
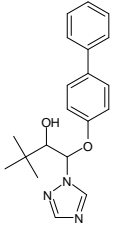
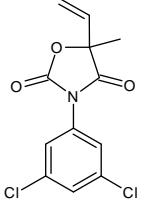
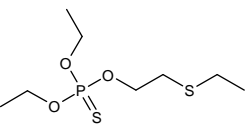
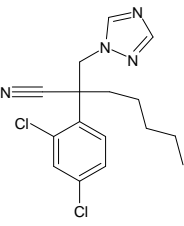
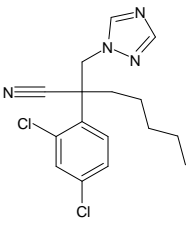
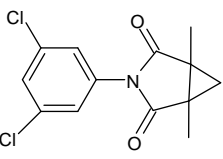
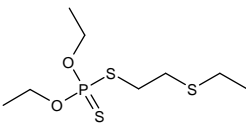
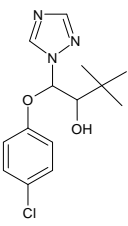
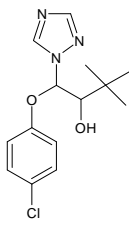
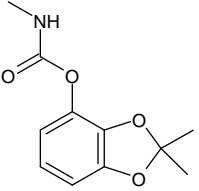
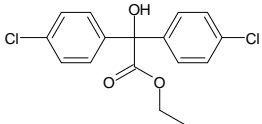
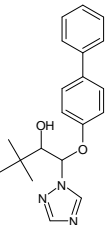
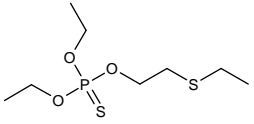
APÊNDICE J – ESTRUTURAS DOS RESULTADOS DA FOSALONA

Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
1) <u>Fosalona</u>	1) Amtriaz	1) Metidation	1) Amtriaz
			
2) Malation	2) Difenzoquat-metil-sulfato	2) Fosalona	2) Procimidona
			
3) Triadimefom	3) Fosalona	3) Propazina	3) Difenzoquat-metil-sulfato
			
4) Dicamba	4) Metidation	4) Ciromazina	4) Propazina
			
5) Fenitroton	5) Alloxydim-sodium	5) Amitraz	5) Fosalona
			

Fonte: Adaptado de NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. PubChem.

Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/>>. Acesso em: 30 jul. 2024.

APÊNDICE K – ESTRUTURAS DOS RESULTADOS DO TRIADIMEFOM

Algoritmo Desenvolvido	KNN	Regressão Logística	Naive-Bayes
1) Triadimefom 	1) Bitertanol 	1) Isofenfos 	1) Isofenfos 
2) Captafol 	2) Isofenfos 	2) Diflubenzurom 	2) Bitertanol 
3) Vinclozolina 	3) Demeton 	3) Miclobutanil 	3) Miclobutanil 
4) Pricimidona 	4) Dissulfotom 	4) Triadimenol 	4) Triadimenol 
5) Bendiocarbe 	5) Clorobenzilato 	5) Bitertanol 	5) Demeton 

Fonte: Adaptado de NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. PubChem.

Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/>>. Acesso em: 30 jul. 2024.