



UFBA

UNIVERSIDADE FEDERAL DA BAHIA
ESCOLA POLITÉCNICA
PROGRAMA DE PÓS GRADUAÇÃO EM
ENGENHARIA INDUSTRIAL - PEI

DOUTORADO EM ENGENHARIA INDUSTRIAL

JÚLIA CAROLINA BRAZ DE FREITAS BIJOS

Predição da ocorrência de fármacos em
efluente de ETE por classificadores
boosting





**UNIVERSIDADE FEDERAL DA BAHIA
ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA INDUSTRIAL**

JÚLIA CAROLINA BRAZ DE FREITAS BIJOS

**PREDIÇÃO DA OCORRÊNCIA DE FÁRMACOS EM EFLUENTE DE ETE
POR CLASSIFICADORES BOOSTING**

Salvador
2025

JÚLIA CAROLINA BRAZ DE FREITAS BIJOS

**PREDIÇÃO DA OCORRÊNCIA DE FÁRMACOS EM EFLUENTE DE ETE
POR CLASSIFICADORES BOOSTING**

Tese apresentada ao Programa de Pós-graduação em Engenharia Industrial, Escola Politécnica, Universidade Federal da Bahia, como requisito ao título de Doutora em Engenharia Industrial.

Área de Concentração: Desenvolvimento Sustentável de Processos e Produtos.

Orientadores: Profa. Dra. Karla Oliveira-Esquerre
Prof. Dr. Luciano Queiroz

Salvador
2025

Ficha catalográfica elaborada pela Biblioteca Bernadete
Sinay Neves, Escola Politécnica - UFBA.

- B594 Bijos, Júlia Carolina Braz de Freitas.
Predição da ocorrência de fármacos em efluentes de ETE por
classificadores boosting /Júlia Carolina Braz de Freitas Bijos. – Salvador,
2025.
155 f.: il. color.
- Orientadora: Profa. Dra. Karla Patrícia Oliveira-Esquerre.
Orientador: Luciano Matos Queiroz.
- Tese (doutorado) – Programa de Pós-graduação em Engenharia
Industrial - Universidade Federal da Bahia - Escola Politécnica, 2025.
- .
1. Ocorrência de Fármacos. 2. Efluentes. 3. Esgoto Sanitário. 4.
Aprendizado de Máquina. I. Oliveira-Esquerre, Karla Patrícia. II.
Queiroz, Luciano Matos. III. Universidade Federal da Bahia. IV. Título.

CDD: 628.2

“PREDIÇÃO DA OCORRÊNCIA DE FÁRMACOS EM EFLUENTE DE ETE POR CLASSIFICADORES BOOSTING”.

JÚLIA CAROLINA BRAZ DE FREITAS BIJOS

Tese submetida ao corpo docente do programa de pós-graduação em Engenharia Industrial da Universidade Federal da Bahia como parte dos requisitos necessários para a obtenção do grau de doutora em Engenharia Industrial.

Examinada por:

Documento assinado digitalmente
gov.br KARLA PATRICIA SANTOS OLIVEIRA RODRIGUEZ
Data: 12/02/2025 12:10:10-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Karla Patrícia Oliveira-Esquerre _____
Doutora em Engenharia Química, Unicamp, Brasil, 2003

Documento assinado digitalmente
gov.br LUCIANO MATOS QUEIROZ
Data: 19/02/2025 13:57:23-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Luciano Matos Queiroz _____
Doutor em Engenharia Civil, Escola Politécnica da Universidade de São Paulo, Brasil, 2009

Profa. Dra. Ana Isabel Carvalho *Ana Carvalho* _____
Doutora em Engenharia Química, Instituto Superior Técnico, Portugal, 2009

Documento assinado digitalmente
gov.br ARIUSKA KARLA BARBOSA AMORIM
Data: 18/02/2025 12:30:22-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Ariuska Karla Barbosa Amorim _____
Doutora em Engenharia Civil, Escola de Engenharia de São Carlos, Universidade de São Paulo, Brasil, 2000

Prof. Dr. Enrique Lopez Droguett *SLM* _____
Doutor em Engenharia de Confiabilidade, University of Maryland at College Park, EUA, 1999

Prof. Dr. Idelfonso Bessa dos Reis Nogueira *IBRN* _____
Doutor em Engenharia Química e Biológica, Universidade do Porto, Portugal, 2018

Documento assinado digitalmente
gov.br SILVIA MARIA ALVES CORREA OLIVEIRA
Data: 14/02/2025 09:58:20-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Silvia Maria Alves Correa Oliveira _____
Doutora em Saneamento, Meio Ambiente e Recursos Hídricos, pela Universidade Federal de Minas Gerais, Brasil, 2006

Salvador, BA - BRASIL
Fevereiro/2025

Aos meu pais, Jane e Rui, que me incentivaram a seguir meus objetivos. Sem vocês eu nada seria.

AGRADECIMENTOS

Para chegar até aqui foi preciso muita determinação e dedicação, fruto não apenas do meu esforço, mas, sobretudo, da convivência com pessoas que me incentivaram e me inspiraram. Agradeço primeiramente a Deus, por ter me direcionado e permitido que os meus planos pudessem se concretizar. Aos meus pais, Jane e Rui, minha irmã Glenda, avó Lacir e tios, em especial tia Débora, meus maiores incentivadores. Ao meu marido, Joel, agradeço pela compreensão nos momentos de ausência e palavras de incentivo diárias. Aos meus amigos, que viveram comigo todos os momentos até aqui e torceram pelo meu sucesso, obrigada.

Agradeço aos meus orientadores, professora Karla e professor Luciano, pela paciência, compreensão e direcionamento, pelo suporte constante ao meu crescimento acadêmico, profissional e pessoal. Encerro esse ciclo tendo vocês como referência para minha vida profissional. Obrigada pelas oportunidades que me apontaram e pela confiança depositada em meu trabalho.

Agradeço ao professor Enrique, que prontamente me recebeu em seu grupo de pesquisa durante o Doutorado Sanduíche e tanto contribuiu para o avanço do meu conhecimento. Sua supervisão foi fundamental para o desenvolvimento dessa pesquisa. Também agradeço a todos que apoiaram a realização dessa etapa: orientadores, professora Viviana e tio Taumaturgo.

Aos amigos do Gamma, que compartilharam desafios e alegrias, obrigada por estarem presentes. Em especial, José Oduque, Robson e estudantes que pude participar da orientação de suas pesquisas, obrigada pelas discussões, risos e parceria.

Aos projetos que atuei, que me proporcionaram ampliar a visão de mundo e me colocaram em uma busca constante por equidade, justiça, avanço social através da ciência e do conhecimento: Meninas na Ciência de Dados e Ciência de dados na Educação Pública, não há palavras que descrevam o quanto minha percepção social foi transformada. A todos os colaboradores do PEI, em especial a Tati, pelo suporte atento e indispensável para sanar dúvidas e fornecer informações quando necessário. Por fim, eu agradeço a todos que, acreditando na educação e ciência, me inspiraram de alguma forma.

O presente trabalho foi realizado com apoio da coordenação de aperfeiçoamento de pessoal de nível superior – brasil (capes) – código de financiamento 001. Este trabalho também contou com o apoio da Fundação de Amparo à Pesquisa do Estado da Bahia – Fapesb – e do programa Capes-Print.

*Para além da curva da estrada
Talvez haja um poço, e talvez um castelo,
E talvez apenas a continuação da estrada.
Não sei, nem pergunto.
Enquanto vou na estrada antes da curva
Só olho para a estrada antes da curva,
Porque não posso ver senão a estrada antes da curva.
De nada me serviria estar olhando para outro lado
E para aquilo que não vejo.
Importemo-nos apenas com o lugar onde estamos.
Há beleza bastante em estar aqui e não noutra parte qualquer.
Se há alguém para além da curva da estrada,
Esses que se preocupem com o que há para além da curva da estrada.
Essa é que é a estrada para eles.
Se nós tivermos que chegar lá, quando lá chegarmos saberemos.
Por ora só sabemos que lá não estamos.
Aqui há só a estrada antes da curva, e antes da curva
Há a estrada sem curva nenhuma.*

*Alberto Caeiro, in "Poemas Inconjuntos" Heterónimo de Fernando
Pessoa*

LISTA DE FIGURAS

Figura 1 – Fontes e rotas de CPEs em matrizes aquáticas	29
Figura 2 – Comparação dos tratamentos aplicados para a remoção de fármacos	44
Figura 3 – Partição em Árvore de Decisão.....	55
Figura 4 -Esquema Ensemble Learning, Boosting	61
Figura 5 – Matriz de Confusão.....	62
Figura 6 – Etapas da metodologia	68
Figura 7 - Etapas desenvolvidas	69
Figura 8 – Pré-processamento de dados de comercialização de fármacos	73
Figura 9 – Variáveis do cálculo de PEC_{inf}	76
Figura 10 - Esquema geral da modelagem de dados	80
Figura 11 - Unidades comercializadas de cada princípio ativo, anualmente.	84
Figura 12 - Unidades comercializadas de Azitromicina, anualmente, 2014 -2020.....	85
Figura 13 - Unidades comercializadas de Amoxicilina anualmente, 2014 - 2020.....	86
Figura 14 - Massa (kg) de princípio ativo comercializada anualmente, 2014 - 2020.	91
Figura 15 - Concentração anual de fármaco - PEC_{inf} (ng. L ⁻¹) encaminhada nos esgotos	93
Figura 16 - Boxplot da Eficiência de Remoção (%) vs Classe terapêutica	95
Figura 17 - Dispersão entre Eficiência de Remoção (%) e Concentração efluente por Classe terapêutica.....	96
Figura 18 - Gráfico de Densidade da Eficiência de Remoção (%) por Classe terapêutica	97
Figura 19 - Curva Precisão-Recall, caso 1000 ng. L ⁻¹	99
Figura 20 - Matriz de Confusão normalizada do modelo XGboost, caso 1.000 ng. L ⁻¹	101
Figura 21 - Curva Precisão-Recall, caso 500 ng. L ⁻¹	102
Figura 22 - Matriz de Confusão normalizada do modelo XGBoost, caso 500 ng. L ⁻¹	103
Figura 23 - Curva Precisão-Recall, caso 100 ng. L ⁻¹	104
Figura 24 - Matriz de Confusão normalizada, Modelo XGBoost, caso 100 ng. L ⁻¹	105
Figura 25 - Summary Plot: Impacto das variáveis na saída do modelo de maior desempenho de cada nível	108
Figura 26 - Desempenho de modelos, caso 1.000 ng. L ⁻¹	112
Figura 27 - Curva Precisão-Recall, caso 500 ng. L ⁻¹	113
Figura 28 - Curva Precisão-Recall, caso 100 ng. L ⁻¹	115
Figura 29 - Impacto das variáveis na resposta do modelo, Antibióticos.....	119

Figura 30 - Impacto médio das variáveis, Análise de Sensibilidade de modelos de Carbamazepina	123
Figura 31 - Impacto médio das variáveis, Análise de Sensibilidade de modelos de Antibióticos.	126
Figura 32 - Conjuntos de tipologias de tratamento de esgoto no Brasil.....	129
Figura 33 - Considerações da pesquisa.	132

LISTA DE QUADROS

Quadro 1 – Classes terapêuticas de fármacos	31
Quadro 2 – Fármacos verificados em matrizes aquáticas	33
Quadro 3 - Pseudoalgoritmo Adaboost	58
Quadro 4 - Pseudoalgoritmo Gradient Boosting Model	59
Quadro 5 – Apresentação de medicamentos	71
Quadro 6 - Fatores de incerteza.....	77

LISTA DE TABELAS

Tabela 1 – Propriedades dos fármacos selecionados para o estudo.....	42
Tabela 2 – Classe prioritária dos Princípios Ativos considerados.....	70
Tabela 3 – Fator de Excreção Humano	75
Tabela 4 – Cálculo de Massa total comercializada (kg) de Carbamazepina, Rio de Janeiro, 2014.	75
Tabela 5 – Dispensação de fármacos por forma de apresentação, acumulada para os anos de 2014 a 2020.	83
Tabela 6 - Forma mais comercializada de cada fármaco e o percentual equivalente nas vendas totais	89
Tabela 7 - Massa de fármacos (kg) comercializada por ano nos municípios representativos, 2014 - 2020.....	90
Tabela 8 - População e vazão afluente de esgoto sanitário–	92
Tabela 9 – Concentração anual afluente de fármaco – PECinf (ng. L ⁻¹).....	92
Tabela 10 - Hiperparâmetros Boosting, modelos de Carbamazepina.....	98
Tabela 11 - Balanceamento de dados, modelos de Carbamazepina	99
Tabela 12 - Avaliação de modelos – 1.000 ng. L ⁻¹	100
Tabela 13 - Métricas XGBoost – 1.000 ng. L ⁻¹	100
Tabela 14 - Avaliação de modelos – 500 ng. L ⁻¹	102
Tabela 15 - Métricas XGBoost – 500 ng. L ⁻¹	103
Tabela 16 - Avaliação de modelos – 100 ng. L ⁻¹	104
Tabela 17 - Métricas XGBoost – 100 ng. L ⁻¹	105
Tabela 18 - Comparação entre modelos em diferentes limites de classificação.....	106
Tabela 19 - Hiperparâmetros Boosting para modelos de Antibióticos.....	110
Tabela 20 - Balanceamento de classes para modelos de Antibióticos.....	111
Tabela 21 - Avaliação de modelos – 1000 ng. L ⁻¹	111
Tabela 22 - Avaliação de modelos – 500 ng. L ⁻¹	112
Tabela 23 - Métricas XGBoost 500 ng. L ⁻¹	114
Tabela 24 - Avaliação de modelos – 100 ng. L ⁻¹	114
Tabela 25 - Métricas XGBoost 100 ng. L ⁻¹	116
Tabela 26 - Comparação entre modelos em diferentes limites de classificação.....	117
Tabela 27 - Vazão de esgoto afluente (m ³) coletado e tratado para análise de sensibilidade	121
Tabela 28 - Acurácia de modelos de Análise de Sensibilidade - Carbamazepina.....	122

Tabela 29 - Acurácia de modelos de Análise de Sensibilidade - Antibióticos.....	124
--	------------

RESUMO

Estações de Tratamento de Esgoto compreendem uma das rotas de emissão de concentrações de fármacos em água superficial, o que pode ofertar riscos a espécies aquáticas e a saúde humana. A partir de dados da quantidade de medicamentos comercializada, é possível estimar as concentrações de fármacos que atingem os cursos d'água após o tratamento realizado pelas estações. Nesse sentido, os métodos de aprendizado de máquina apresentam capacidade de aprendizado e, por isso, podem ser uma ferramenta relevante para a proposição de uma nova abordagem de modelos confiáveis e robustos para a predição da ocorrência de fármacos. Esta pesquisa teve por objetivo geral propor um mecanismo de modelagem de classificação capaz de prever o nível das concentrações de alguns fármacos nos efluentes de estações de tratamento de esgoto. Para tal, foram utilizados dados de comercialização de fármacos para os municípios brasileiros referentes ao período de 2014 a 2020, registros de população e vazão afluente de esgotos municipais e, registros de concentrações de ocorrência de antibióticos, anti-inflamatórios e drogas psiquiátricas em estações de tratamento. O software *RStudio* foi utilizado para a manipulação dos dados de comercialização de fármaco e obtenção da massa anual comercializada, que foi utilizada posteriormente, para o cálculo da concentração afluente (ng. L^{-1}) de fármacos para tratamento. Três métodos de classificação *Boosting* foram implementados na linguagem *Python*, executados para três níveis de rotulação: 100 ng. L^{-1} , 500 ng. L^{-1} e 1000 ng. L^{-1} , a fim de prever a classe da concentração de saída dos fármacos para Carbamazepina e para antibióticos selecionados. Em ambas as abordagens, o modelo XGBoost atingiu o maior desempenho em todos os níveis. Para a abordagem de Carbamazepina os modelos XGBoost atingiram F1-score de 63%, 91% e 90%. Para a abordagem com antibióticos o modelo XGBoost obteve F1-score de 99%, 90% e 85% para os limites 1000 ng. L^{-1} , 500 ng. L^{-1} e 100 ng. L^{-1} , respectivamente. Para ambas as abordagens foi possível obter modelos de desempenho satisfatório. No entanto, o limite de corte de 1000 ng. L^{-1} demonstrou ser o mais desafiador, dado que gerou amostras de teste com classes altamente desbalanceadas, enquanto os limites de corte de 500 ng. L^{-1} e 100 ng. L^{-1} possibilitaram a construção de modelos mais viáveis. Os ajustes construídos revelaram que o uso dos classificadores *Boosting* apresenta potencial como alternativa de complementação ao monitoramento e controle de fármacos no esgoto sanitário. Ademais, a utilização de dados de acesso livre disponibilizados pela ANVISA pode fornecer uma reflexão dos potenciais e das limitações para futuros estudos, visto que até o momento ainda não houve uma ampla exploração desses dados em pesquisas. Adicionalmente, a base de dados de ocorrência de fármacos em ETE, construída para esta pesquisa, poderá compor as atividades investigativas de outros pesquisadores, a partir da agregação de novos dados, verificação de outras técnicas de modelagem ou realização de outros tipos de análises.

Palavras-chave: Ocorrência de fármacos. Efluentes. Esgoto Sanitário. Aprendizado de Máquina.

ABSTRACT

Wastewater Treatment Plants are a significant pathway for the emission of pharmaceutical concentrations into surface water, which can pose risks to aquatic species and human health. Using data on the quantity of medicines sold, it is possible to estimate the concentrations of pharmaceuticals that reach the surface waters after the treatment carried out in the treatment plants. In this sense, machine learning methods show learning capacity and, therefore, can be a relevant tool for proposing a new approach to reliable and robust models for predicting the occurrence of medicines. The research proposed a classification modeling mechanism capable of predicting the concentration level of some pharmaceuticals in effluents from wastewater treatment plants. To this end, drug sales data were used for Brazilian municipalities from 2014 to 2020, population and influent flow of municipal wastewater treatment plants, and data of occurrence concentrations of antibiotics, anti-inflammatories, and psychiatric drugs in treatment plants. RStudio software was used to manipulate drug sales data and obtain the annual mass sold, which was later used to calculate the influent concentration (ng. L^{-1}) of drugs for treatment. Three Boosting classification methods were implemented in the Python language, running for three levels: 100 ng. L^{-1} , 500 ng. L^{-1} and 1000 ng. L^{-1} , to predict the pharmaceutical outlet concentration class for Carbamazepine and selected antibiotics. In both approaches, the XGBoost model achieved the highest performance at all levels. For the Carbamazepine approach, the XGBoost models achieved F1 scores of 63%, 91%, and 90%. For the antibiotic approach, the XGBoost model obtained an F1-score of 99%, 90%, and 85% for the 1000 ng. L^{-1} , 500 ng. L^{-1} and 100 ng. L^{-1} , respectively. For both approaches, it was possible to obtain models with satisfactory performance. However, the cutoff limit of 1000 ng. L^{-1} proved to be challenging, as it generated samples with highly imbalanced classes, while the cutoff limits were 500 ng. L^{-1} and 100 ng. L^{-1} enabled the construction of more viable models. The models revealed that Boosting classifiers have potential as an alternative for monitoring and controlling pharmaceuticals in sanitary sewage. The use of open-access data made available by ANVISA reflects the potentials and limitations for future studies since to date there has not been a broad exploration in research. Additionally, the drug occurrence database in WWTPs, built for this research, may be part of the investigative activities of other researchers, by aggregating new data, verifying other modeling techniques, or carrying out other types of analyses.

Keywords: Pharmaceutical occurrence. Effluents. Sanitary sewage. Machine Learning.

SUMÁRIO

1	INTRODUÇÃO E CONTEXTUALIZAÇÃO	19
2	OBJETIVOS	23
2.1	ESTRUTURA DO DOCUMENTO	23
3	REVISÃO DE LITERATURA.....	25
3.1	COMERCIALIZAÇÃO DE FÁRMACOS	25
3.2	CONTAMINANTES EMERGENTES NO MEIO AMBIENTE: A OCORRÊNCIA DE FÁRMACOS	27
3.2.1	Ocorrência de fármacos em matrizes aquáticas e aspectos limitantes do monitoramento	31
3.2.2	Ocorrência de fármacos em ETE	36
3.2.3	Remoção de fármacos em ETE.....	38
3.2.4	Marcos Regulatórios.....	44
3.3	O USO DE MODELAGEM EM SISTEMAS DE REMOÇÃO DE FÁRMACOS	47
3.4	APRENDIZADO DE MÁQUINA	49
3.3.1	Métodos de Aprendizado de Máquina no controle e monitoramento de contaminantes emergentes	51
3.3.2	Métodos de classificação baseados em árvores	54
3.3.3	Métricas de avaliação de modelos de classificação	62
3.3.4	Explicabilidade de modelos de aprendizagem: método SHAP	63
3.5	INOVAÇÃO DA PROPOSTA	65
4	METODOLOGIA	66
4.1	DESCRIÇÃO GERAL DA PESQUISA	66
4.2	CONSTRUÇÃO DA BASE DE DADOS PARA O MODELO	69

4.2.1. Seleção de municípios representativos e fármacos comercializados - ANVISA ...	69
4.2.2. Pré-processamento de dados – ANVISA	70
4.2.3. Estimativa de PEC_{inf} com base na comercialização de fármacos	73
4.2.4. Coleta de concentrações de ocorrência de fármacos nos esgotos sanitários.....	76
4.2.5. Caracterização dos dados.....	77
4.3 IMPLEMENTAÇÃO DA MODELAGEM DE DADOS	79
4.3.1 Modelagem proposta para Carbamazepina	80
4.3.2 Modelagem proposta para os Antibióticos	81
4.3.3 Análise de Sensibilidade	82
5 RESULTADOS E DISCUSSÃO	83
5.1 DADOS ANVISA – COMERCIALIZAÇÃO ANUAL DE FÁRMACOS	83
5.2 CONCENTRAÇÃO AFLUENTE (PEC_{INF}) PARA OS MUNICÍPIOS	92
5.3 DADOS DA LITERATURA – CONCENTRAÇÕES DE OCORRÊNCIA EM ESGOTOS SANITÁRIOS	94
5.4 IMPLEMENTAÇÃO DOS MODELOS	97
5.4.1 Modelos para a Carbamazepina.....	98
5.4.2 Modelos para os Antibióticos.....	110
5.5 ANÁLISE DE SENSIBILIDADE.....	121
5.5.1 Carbamazepina	121
5.5.2 Antibióticos.....	124
5.6 RELEVÂNCIA DA ABORDAGEM NO CONTEXTO AMBIENTAL	127
6 CONCLUSÕES	133
7 SUGESTÕES PARA TRABALHOS FUTUROS	135
REFERÊNCIAS	136

APÊNDICE A – AVALIAÇÃO DOS MODELOS.....	152
ANEXO A – DICIONÁRIO DE DADOS ABERTOS DA ANVISA.....	155

1 INTRODUÇÃO E CONTEXTUALIZAÇÃO

Estima-se que o faturamento global da indústria de fármacos em 2023 foi de, aproximadamente, US\$ 1,6 trilhões. Liderado pelos Estados Unidos e acompanhado por países em desenvolvimento como Brasil, Colômbia, Índia e Egito, o consumo de fármacos aumentou consideravelmente nos últimos 20 anos, com crescimento anual equivalente a 5,8% a partir do ano de 2017 (MIKULIC, 2023; GONZÁLEZ PEÑA; LÓPEZ ZAVALA; CABRAL RUELAS, 2021).

Os fármacos são parte de uma ampla classe de contaminantes de preocupação emergente (CPEs) para os quais ainda não há conhecimento consolidado sobre os efeitos negativos do descarte sobre os ecossistemas ou a saúde humana. Parte da concentração dos fármacos consumida é excretada pelo corpo humano após a metabolização, o que confirma a hipótese que a inserção dessas substâncias no ambiente aquático continuará ocorrendo de forma ininterrupta e aumentada por meio do lançamento de esgotos tratados, ou não, nesse compartimento ambiental (PEREIRA *et al.*, 2020).

Esse aspecto é consequência do fato de que os arranjos tecnológicos aplicados em estações de tratamento de esgotos (ETE) não são suficientes para a completa remoção desses compostos. Portanto, considerando o cenário de atendimento pleno com sistemas de esgotamento sanitários, pode-se afirmar que as ETE são a principal fonte de emissão de concentrações de fármacos em cursos hídricos. O problema agrava-se porque muitos cursos d'água receptores de esgotos também são mananciais para produção de água para abastecimento público (MONTAGNER *et al.*, 2019; SOARES; SOUZA, 2020).

Ademais, deve-se considerar a estimativa que mostra que, aproximadamente, 4 bilhões de pessoas não têm acesso a qualquer serviço de saneamento básico, o que aumenta a probabilidade de que fármacos excretados nos esgotos domésticos ou descartados inadequadamente com resíduos sólidos domiciliares contaminem o meio ambiente (SOARES; SOUZA, 2020; UNESCO, 2023).

Portanto, o contato direto com o esgoto não tratado agrava os riscos associados à ocorrência de fármacos. Essa situação pode ser constatada, sobretudo, em países de renda baixa a média que apresentam infraestrutura limitada para o gerenciamento de esgoto e, onde já foram constatadas altas concentrações de fármacos em águas superficiais, por exemplo (WILKINSON *et al.*, 2022).

O monitoramento da ocorrência de fármacos nas diferentes matrizes ambientais é realizado por meio de campanhas amostrais e custosas análises laboratoriais. Entretanto, devido

aos elevados custos financeiros e limitações tecnológicas, apenas uma ínfima parcela desses compostos está representada por meio de dados de monitoramento, geralmente, esparsos e distribuídos desigualmente entre os países e continentes (ACUÑA et al., 2020; WILKINSON et al., 2022).

Assim, não há um processo bem estabelecido para coleta, análise e divulgação de dados de ocorrência dos compostos e, diante da alta quantidade de fármacos consumida e complexidade molecular, torna-se inviável rastrear todos por análises experimentais. Por isso, é fundamental desenvolver ferramentas capazes de auxiliar a identificação e quantificação desses contaminantes nos efluentes de forma assertiva e ágil (CHEN et al., 2025; KOSTICH; BATT; LAZORCHAK, 2014).

Uma abordagem convencionalmente aplicada para estimar a Concentração Ambiental Predita (*Predicted Environmental Concentration* – PEC) de fármacos em esgotos brutos afluentes às ETE, se baseia na quantidade de princípio ativo comercializado, volume de esgoto gerado, população local e percentual do ativo excretado pelo metabolismo humano (HE et al., 2020; VERLICCHI et al., 2014).

Entretanto, essa abordagem ainda depende extensivamente de dados experimentais da ocorrência de fármacos em ETE, seu potencial de remoção e, negligencia diferentes relações entre entradas e saídas. Dado que os poluentes em ETE refletem condições dinâmicas, não lineares e variedade de parâmetros, a estimativa da ocorrência desses compostos pode ser melhorada através de modelos que considerem essas condições.

Diante desse cenário, a modelagem matemática que resulta na sistematização de dados e informações em bases acessíveis pode possibilitar a ampliação da rastreabilidade dessas substâncias no meio ambiente usando menos recursos. O aumento do poder computacional associado a novas possibilidades de obtenção de dados consolidou os métodos de aprendizado de máquina como um caminho versátil para explorar problemas por meio de técnicas de agrupamento, predição e classificação (TAOUFIK *et al.*, 2022).

Assim, através dos padrões dos dados de entrada, identifica-se a relação entre as variáveis e encontra-se o melhor ajuste de parâmetros para o método utilizado. Essa capacidade de aprendizado permite que novos dados sejam inseridos e de acordo com o novo perfil identificado, o método é novamente ajustado, o que demonstra capacidade de generalização.

Nesse sentido, esses métodos possuem capacidade de lidar com dados desafiadores que podem apresentar natureza não linear e instâncias incompletas, e, dentre as opções, há métodos que não fazem pressuposições sobre os dados. Mesmo nessas condições, é possível extrair

conhecimento acerca das variáveis e mapear com precisão as relações entre entradas e saídas (FAN *et al.*, 2018).

Diferentes métodos têm sido aplicados no monitoramento e controle da ocorrência de fármacos e contaminantes emergentes em matrizes aquáticas ou efluentes. Aplicações recentes destacam a identificação do perfil temporal da concentração de contaminantes por meio de modelo Rede de Memória de Curto Prazo (WU; LI; XIONG, 2021), uso dos modelos Redes Neurais Artificiais, Floresta Aleatória, Árvore de Decisão e Máquina de Vetores de Suporte para predição da remoção de fármacos em efluentes (TAOUFIK *et al.*, 2022) e, ainda, Análise de Agrupamento para identificar fatores mais influentes na eficiência de remoção ou para o reconhecimento de novos contaminantes (VAKARELSKA *et al.*, 2022; WANG; FENNER; HELBLING, 2020).

Técnicas de aprendizado pela abordagem *boosting* como XGBoost, LightGBM e Adaboost aplicadas a modelos de árvore têm demonstrado desempenho consideravelmente superior às técnicas convencionais para a predição da qualidade de efluentes. Em geral, os modelos consideram como preditores pH, temperatura, oxigênio dissolvido, teor de metais tóxicos, entre outros parâmetros operacionais (NGUYEN *et al.*, 2024). Outras abordagens incluem mais variáveis relevantes para o problema, como parâmetros físico-químicos, dados de toxicidade, além da concentração de alguns poluentes emergentes, incluindo fármacos (CHEN *et al.*, 2025; WANG *et al.*, 2022).

As recentes aplicações de modelos *boosting* indicam que o aprendizado lento a partir da informação residual das árvores, capaz de promover a melhoria da capacidade preditiva e robustez ao *overfitting*, têm se mostrado extremamente vantajoso diante da capacidade de apontar variáveis e compostos prioritários ao monitoramento em ETE.

Em geral, os métodos citados permitem utilizar dados categóricos ou numéricos, lidam bem com valores faltantes e valores aberrantes, são escaláveis, e apresentam potencial de interpretabilidade (HASTIE; FRIEDMAN; TIBSHIRANI, 2001). No entanto, os estudos que aplicaram métodos de Aprendizado de Máquina até o presente momento, ainda são escassos e não consideram a comercialização de fármacos como importante variável de interesse.

Desse modo, a literatura aponta que (I) restrições à ampla coleta de dados de ocorrência ambiental de fármacos evidenciam a importância do uso de modelos preditivos, (II) a natureza complexa dos dados exige o uso de modelos com alta capacidade de aprendizado e, (III) o aumento na comercialização de fármacos confirma a necessidade de se considerar variáveis desse grupo como preditores. Porém, a consideração de dados da comercialização em conjunto

com variáveis relacionadas ao processo de remoção empregado em ETE ainda não foi identificada na literatura científica.

O contexto apresentado revela que a implementação de métodos de Aprendizado de Máquina pode apresentar potencial para gerar modelos de alto desempenho capazes de prever a concentração final de fármacos após o tratamento nas ETE, através da associação de dados prévios de venda de fármacos e concentração de ocorrência desses compostos em efluente de ETE. Registros de comercialização de fármacos apontam informações relevantes sobre o perfil de consumo da população e por isso são variáveis independentes para estimar a concentração de entrada (afluente) de fármacos em ETE, ao passo que variáveis do processo de tratamento possibilitam estimar a concentração desses compostos no efluente tratado (VERLICCHI *et al.*, 2014).

A pesquisa proposta tem potencial de lograr resultados que poderão auxiliar agentes públicos na definição de futuras regulamentações rumo à priorização para monitoramento e controle de fármacos, bem como traçar estratégias de remediação, controle, e redução dos impactos negativos observados. Conjuntamente, torna-se um mecanismo em prol da garantia da disponibilidade e gestão sustentável da água potável e do saneamento para todos, Objetivo do Desenvolvimento Sustentável 6 (ODS 6) e, conseqüentemente, manutenção da saúde pública (ONU, 2015).

Indiretamente, esta pesquisa favorece o desenvolvimento do ODS 10, ODS 11, ODS 12, ODS 13, ODS 14, ODS 15, que versam sobre Redução das Desigualdades; Cidades e Comunidades Sustentáveis; Consumo e Produção Responsáveis; Ação contra a mudança global do clima; Vida na água e, Vida Terrestre, respectivamente.

Portanto, a presente pesquisa contribui com uma nova abordagem que estabelece um mecanismo de classificação baseado em árvores de decisão no contexto *boosting*, para predição da concentração efluente de fármacos em ETE a partir de dados de comercialização de medicamentos. O estudo tem capacidade para complementar o processo convencional de monitoramento, se adaptar a diferentes cenários e mitigar a necessidade de executar custosas campanhas para coletas de amostras.

Nesse sentido, a proposição desta pesquisa parte da pergunta norteadora “É possível obter uma ferramenta de predição da concentração de fármacos encaminhada aos cursos d’água receptores de esgotos tratados por meio de classificadores baseados em árvores?”, de modo que a hipótese a ser testada é a de que é possível aplicar essas técnicas para propor um modelo preditivo adequado, com possibilidade de complementar o monitoramento de fármacos no ambiente.

2 OBJETIVOS

Esta pesquisa tem como objetivo geral propor um mecanismo de modelagem de classificação capaz de prever o nível das concentrações de alguns fármacos nos efluentes de estações de tratamento de esgoto (ETE).

Para tanto, foram definidos os seguintes objetivos específicos:

- Determinar fármacos de maior relevância presente nos esgotos com base em estudos da literatura;
- Verificar a capacidade de remoção desses compostos nas ETE a partir de valores de eficiência de remoção provenientes de diferentes tecnologias de tratamento e obtidos em trabalhos da literatura científica;
- Propor modelos de predição e avaliar o desempenho obtido.

2.1 ESTRUTURA DO DOCUMENTO

Este documento encontra-se dividido em capítulos de acordo com a estrutura proposta pelo Manual de Estilo Acadêmico, 6ª edição (LUBISCO; VIEIRA, 2019) e são respeitadas as normas do Programa de Pós-Graduação em Engenharia Industrial da Universidade Federal da Bahia (PEI – UFBA).

O Capítulo 1 apresenta a descrição da natureza do problema investigado, abordagem metodológica adotada, motivação e justificativa, hipótese considerada, contribuição para o estado da arte e o potencial de aplicação dos modelos construídos.

O Capítulo 2 aponta o objetivo geral e específicos, e informa ao leitor o modo como os capítulos deste documento estão organizados.

O Capítulo 3 apresenta a Revisão de Literatura do tema, estruturada em quatro sessões principais. Na primeira, buscou-se apresentar o cenário da comercialização de medicamentos de forma a mostrar o impacto e importância dessa indústria em uma perspectiva global. Em seguida, buscou-se apresentar informações sobre a ocorrência de contaminantes de preocupação emergente em matrizes aquáticas, o monitoramento e remoção de fármacos em ETE e um panorama das normativas de controle propostas por órgãos internacionais. Na terceira sessão, apresenta-se características que tornam desafiador o processo de remoção de poluentes em ETE. Além disso, apresenta-se os princípios de modelos relevantes baseados em uma abordagem mecanicista ou molecular. Na quarta sessão, os conceitos fundamentais de Aprendizado de Máquina são apresentados e complementa-se com a exposição de aplicações recentes que

demonstram o potencial dos métodos na quantificação e remoção de contaminantes. Na quinta sessão, reforça-se o caráter inovador desta pesquisa.

O Capítulo 4 se subdivide em três sessões principais para melhor delimitar os métodos realizados. A primeira sessão retoma a especificação do problema de pesquisa, como foi abordado na presente investigação e, ainda, apresenta uma descrição das fontes de dados utilizadas na pesquisa. Na segunda sessão, apresenta-se os critérios para a seleção de municípios e fármacos considerados, a coleta de dados, seleção de variáveis, filtros aplicados e demais pré-tratamentos realizados nas bases de dados iniciais provenientes da Agência Nacional de Vigilância Sanitária (ANVISA) e da Agência Nacional de Águas e Saneamento Básico (ANA). Também é apresentado o processo de coleta e organização dos registros da ocorrência de fármacos em ETE a partir da literatura, essenciais para a construção da base de dados final utilizada para a modelagem. Por fim, uma caracterização qualitativa é apresentada com o objetivo de informar sobre a instauração, manutenção e aplicação das bases de dados utilizadas, para que o leitor conheça possíveis fatores de imprecisão. Na terceira sessão, são incluídas as considerações acerca da predefinição de classes dos modelos, as bibliotecas utilizadas para implementação dos métodos e o detalhamento dos procedimentos da Análise de Sensibilidade.

Na primeira sessão do Capítulo 5, apresenta-se os resultados verificados para a comercialização anual de fármacos segundo a base da ANVISA. Na segunda sessão, apresenta-se os resultados do procedimento de cálculo de concentração afluente de fármacos em ETE. Na terceira sessão, são mostradas informações sobre os registros de ocorrência de fármacos coletados na literatura. Na quarta sessão, são apresentados os resultados dos modelos construídos. Na quinta sessão, apresenta-se os resultados verificados para a Análise de Sensibilidade. E, na sexta sessão, é apresentada uma discussão sobre os aspectos ambientais que evidenciam a importância do mecanismo proposto.

O Capítulo 6 relata as conclusões obtidas, reforçando a contribuição da pesquisa e, aponta como a abordagem desenvolvida pode melhorar o estado da arte.

O Capítulo 7 sugere as possibilidades futuras de investigações no tema.

3 REVISÃO DE LITERATURA

3.1 COMERCIALIZAÇÃO DE FÁRMACOS

A Lei nº 5.991, de 17 de dezembro de 1973, define o termo “medicamento” como: “produto farmacêutico, tecnicamente obtido ou elaborado, com finalidade profilática, curativa, paliativa ou para fins de diagnóstico” (BRASIL, 1973, p.1). Assim, o medicamento, obrigatoriamente, apresenta forma e formulação farmacêutica que atendam aos requisitos de qualidade e segurança, além de eficácia cientificamente comprovada para ser comercializado.

Os medicamentos são produtos farmacêuticos que contém em sua formulação um ou mais fármacos combinados, também conhecidos como “Insumo Farmacêutico Ativo (IFA)” ou “Princípio Ativo” (AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2009). Outras terminologias amplamente utilizadas são os correspondentes em inglês *Active Pharmaceutical Ingredients (APIs)* e *Pharmaceuticals Active Compounds (PhACs)*, sendo o último termo amplamente utilizado em investigações relacionadas à ocorrência desses compostos no meio ambiente (KOSTICH; BATT; LAZORCHAK, 2014; PARIDA *et al.*, 2021; SANTOS *et al.*, 2020).

Uma tendência global de crescimento na comercialização de produtos farmacêuticos tem sido constatada nos últimos anos, de modo que a receita gerada por este setor no ano de 2023 foi de 1,6 trilhões de dólares, o que demonstra aumento considerável em comparação ao ano de 2001, que apresentou valor de 390,2 milhões de dólares. Os Estados Unidos figuram como o país líder, representa 42,6% dos gastos do mercado farmacêutico global e apresenta maior gasto farmacêutico per capita, um valor médio de US\$ 1.310 dólares por pessoa (MIKULIC, 2023b).

O crescimento dos mercados da China e Índia também é observado e se deve ao fato de que esses países apresentaram aumento do Produto Interno Bruto (PIB), ao passo que vivenciam o envelhecimento da população atrelado ao sedentarismo e níveis de poluição urbana que ocasionam problemas respiratórios (GONZÁLEZ PEÑA; LÓPEZ ZAVALA; CABRAL RUELAS, 2021).

Com uma participação de 1,9%, o Brasil ocupa a décima posição e figura como único país da América Latina entre os dez maiores mercados farmacêuticos mundiais (MIKULIC, 2023b). Projeções atualizadas da companhia *Quintiles and IMS Health, Inc.* (IQVIA, 2023) apontam que até 2027 o mercado farmacêutico da região irá aumentar em 1,6%. Brasil e México

lideram como potências para o crescimento, de forma que juntos poderão representar 2/3 do gasto total da região.

Quando se realiza uma análise preliminar, o crescimento projetado poderia ser atribuído ao aumento populacional estimado para a América Latina (IQVIA, 2023). Entretanto, a possibilidade de compra de medicamentos sem a necessidade de prescrição médica pelo usuário final associada ao favorecimento da entrada de novos fármacos no mercado, também podem contribuir para o agravamento do cenário (GONZÁLEZ PEÑA; LÓPEZ ZAVALA; CABRAL RUELAS, 2021).

A ampliação do acesso a medicamentos é parte do Objetivo de Desenvolvimento Sustentável (ODS) 3 da Agenda 2030 da ONU, que versa sobre a garantia do acesso à saúde de qualidade e promoção do bem-estar para todos. Especificamente, uma das metas desse ODS visa universalizar a saúde para garantir de forma acessível e segura o acesso a serviços essenciais de saúde, medicamentos e vacinas (ONU, 2015).

O acesso a medicamentos também é promovido pela Organização Mundial de Saúde (OMS) através da definição de uma lista de medicamentos essenciais, que desde a sua concepção apresenta tendência de aumento. No ano de 1987, 37% da população mundial não tinha acesso regular aos medicamentos listados. Já em 1999, esse valor foi reduzido para cerca de 1,3 a 2,1 bilhões de pessoas, correspondente a 30% da população mundial. Vale ressaltar que nessa situação o continente africano e a Índia eram as regiões com menor cobertura de acesso (WORLD HEALTH ORGANIZATION, 2004). A expectativa é de que até o ano de 2030 a provisão de medicamentos essenciais continue aumentando (PIGGOTT *et al.*, 2022).

Nesse sentido, os esforços realizados têm logrado o aumento da cobertura de acesso a medicamentos como no caso da Índia, onde, em 2012, foi implementado um programa para viabilizar e reconhecer a produção de medicamentos por empresas de menor escala e, conseqüentemente, provocou a redução de preços do mercado devido à concorrência (WORLD HEALTH ORGANIZATION, 2017).

No Brasil, a Política Nacional de Medicamentos estabelecida pela Portaria nº 3.916 do Ministério da Saúde (1998) tem exercido um papel fundamental no acesso à medicamentos de forma segura, eficaz, com garantia de qualidade, baixo custo e promovendo o uso racional. Aliada à Política Nacional de Assistência Farmacêutica, o desenvolvimento das iniciativas alcançou resultados notórios com a implementação de projetos como Medicamentos genéricos (Lei nº 9.787, de 10 de janeiro de 1999) e a Farmácia Popular do Brasil (2021) (BRASIL, 2004).

González Peña *et al.* (2021), destacam que em nível mundial a comercialização dos fármacos oncológicos, antidiabéticos, de doenças respiratórias e autoimunes, antibióticos e

princípios ativos de vacinas são os que tem gerado maior receita no setor. Os autores ainda apontam que os fármacos neurológicos ou psiquiátricos, em conjunto com os cardiovasculares e oncológicos, estão relacionados as causas de morte mais típicas da população, o que reforça a importância de se investigar essas classes terapêuticas.

Verifica-se que, no ano de 2018, um total de 150 milhões de embalagens de antidepressivos, pertencentes a classe terapêutica de fármacos psiquiátricos, foram vendidas no Brasil (PIVETTA *et al.*, 2020). Atualmente, essa classe ocupa o nono lugar no ranking das dez classes terapêuticas com maior faturamento de vendas no país, o que reforça o padrão observado mundialmente (ANVISA, 2023). Ressalta-se que os antibióticos também são foco de interesse em pesquisas devido à sua ampla comercialização e impactos negativos que podem ocasionar ao meio ambiente e à saúde humana como a resistência antimicrobiana.

Diante disso, constata-se que o uso de fármacos em prol da melhoria da qualidade de vida e aumento da expectativa de vida do usuário final tende a aumentar, apoiado em uma ampla estrutura que envolve a atuação de rentáveis indústrias farmacêuticas, organizações mundiais e políticas públicas. Porém, diante da eficácia cientificamente comprovada para tratar uma enfermidade específica, não há um consenso acerca da obrigatoriedade de uma avaliação de risco devido à exposição ao fármaco em matrizes ambientais (GONZÁLEZ PEÑA; LÓPEZ ZAVALA; CABRAL RUELAS, 2021).

Verifica-se um desalinhamento de expectativas, já que o sucesso da ampliação de acesso aos medicamentos proposta pelo ODS 3 – “Saúde e bem estar”, medida extremamente necessária, interfere em algum nível no cumprimento do ODS 6 - “Água Potável e Saneamento”, que visa a disponibilização e gestão da água potável e saneamento para todas as pessoas. Portanto, pode-se afirmar que não é prioridade do mercado farmacêutico avaliar o impacto ambiental pela ocorrência de fármacos e que esse critério não é impeditivo para a comercialização desses componentes e, por conseguinte, a regulação da disposição final dessas substâncias.

3.2 CONTAMINANTES EMERGENTES NO MEIO AMBIENTE: A OCORRÊNCIA DE FÁRMACOS

Os Contaminantes de Preocupação Emergente (CPEs), Contaminantes Emergentes (CEs) ou microcontaminantes, correspondem a uma vasta classe de compostos de origem natural ou sintética que ocorrem nas matrizes ambientais como resultado de processos naturais

ou da atividade antropogênica (resultados de atividades humanas como pecuária, por exemplo). São também identificados pelo termo em inglês *Contaminants of Emerging Concern (CECs)*.

No Brasil, esses compostos não obedecem a padrões regulatórios, o que significa que não integram rotinas estruturadas de monitoramento no meio ambiente. Embora pouco se compreenda sobre os mecanismos de degradação e destino dos CPEs no meio ambiente, pesquisas apontaram que a presença nos compartimentos ambientais está associada a efeitos nocivos ao meio ambiente ou à saúde humana, ainda que ocorram em concentrações traço, em nível de ng. L⁻¹ ou µg. L⁻¹ (EPA, 2008; GEISSEN *et al.*, 2015; PARIDA *et al.*, 2021).

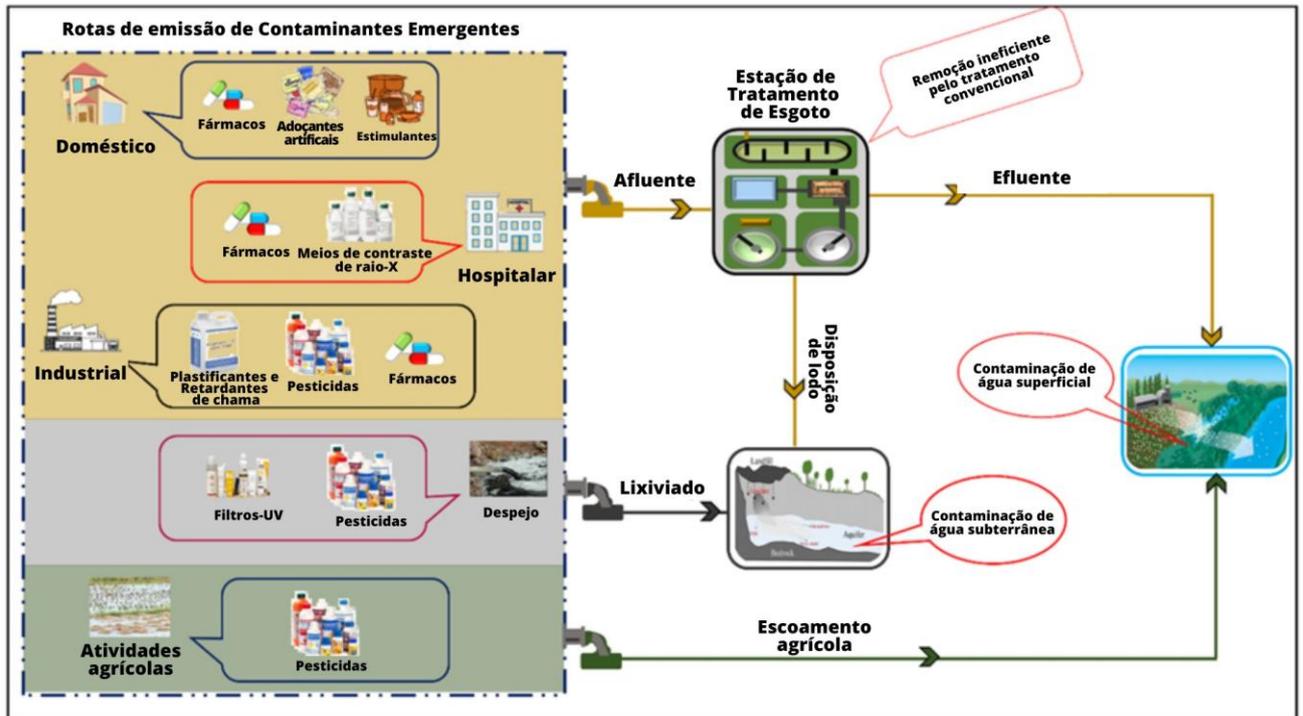
A classe de CPEs engloba centenas de novos compostos e metabólitos ou químicos já existentes, cuja disponibilidade no ambiente foi influenciada pelo crescimento industrial e aumento de demanda por produtos industrializados. Esses compostos passaram a ser investigados no final dos anos 1970, mas, foi a partir de 2012, que se verificou um aumento expressivo na quantidade de publicações sobre o tema em decorrência, principalmente, de avanços nas técnicas analíticas como cromatografia gasosa-espectrometria de massa (GC-MS/MS), cromatografia líquida-espectrometria de massa (LC-MS/MS), cromatografia líquida de ultra eficiência acoplado à espectrometria de massas (UPLC/MS), utilizados para a identificação e quantificação dessas substâncias (EPA, 2008; LIMA *et al.*, 2017a; PARIDA *et al.*, 2021).

Participam dessa classe de contaminantes os Fármacos e Produtos de Cuidado Pessoal (*Pharmaceuticals and Personal Care Products, PPCPs*), Disruptores ou Perturbadores Endócrinos (*Endocrine disrupting chemicals, EDCs*) como hormônios sintéticos e naturais, além de Pesticidas, Herbicidas e Agrotóxicos, Retardantes de chama (*Polybrominated Flame Retardants*), Compostos Perfluorados (*Perfluorinated compounds, PFCs*), Nanomateriais, Corantes, Drogas ilícitas, Adoçantes artificiais, Microplásticos, Metais tóxicos, Meios de contraste de raio-X, Estimulantes, Dioxinas, Siloxanos ou outros, podendo haver sobreposição entre as classes (MONTAGNER; VIDAL; ACAYABA, 2017; PESCARA, 2014).

Os CPEs atingem as matrizes ambientais por meio do descarte em diferentes vias de contaminação como descarte incorreto de medicamentos e produtos de cuidado pessoal nos resíduos sólidos domiciliares, aplicação de compostos na agricultura, consumo de medicamentos, adoçantes ou estimulantes pela população e posterior excreção dos resíduos no esgoto doméstico, uso de substâncias em processos industriais, entre outros. Por essas vias de contaminação, os CPEs atingem solos e, as matrizes aquáticas de águas subterrâneas, esgotos industriais e domésticos e o efluente do tratamento, águas superficiais, águas tratadas e até

mesmo água para consumo humano (PESCARA, 2014). A Figura 1 exibe as principais rotas de inserção dos CPEs no meio ambiente.

Figura 1 – Fontes e rotas de CPEs em matrizes aquáticas



Fonte: Traduzido de Parida *et al.* (2021).

A ocorrência dos CPEs em matrizes aquáticas é particularmente preocupante, uma vez que esses compostos apresentam baixa degradabilidade. Como consequência, permanecem no ambiente por tempo indefinido e, portanto, podem provocar alterações no ecossistema aquático e elevar os riscos de exposição de seres humanos por meio da captação de água para abastecimento municipal contendo concentrações traços desses compostos.

Atenção especial tem sido prestada ao grupo dos fármacos, que somam cerca de 3.000 compostos utilizados na produção de medicamentos humanos e veterinários, para comercialização com ou sem prescrição médica. Dentre a enorme quantidade catalogada, apenas uma pequena parcela é de fato investigada e, frente ao alto consumo pela população, a ocorrência de compostos dessa classe se tornou foco de pesquisas (RICHARDSON; TERNES, 2014).

Os fármacos são persistentes no ambiente; apresentam potencial de bioacumulação; são resistentes ao tratamento convencional aplicado em ETE e estações de tratamento de água (ETA); apresentam potencial de toxicidade e podem gerar metabólitos também tóxicos,

ofertando riscos ao ecossistema ou aos seres vivos (LIMA *et al.*, 2017a; MONTAGNER; VIDAL; ACAYABA, 2017; PARIDA *et al.*, 2021).

O risco ocasionado aos organismos aquáticos frente à exposição a esses compostos é avaliado por ensaios ecotoxicológicos por meio de técnicas como a avaliação de risco ambiental (*Environmental Risk Assessment, ERA*), que verifica o potencial de toxicidade por meio de ensaios *in vivo* em espécies representativas, expostas à diferentes concentrações do contaminante por um tempo determinado. Posteriormente, o risco ambiental é verificado através do cálculo do Coeficiente de Risco (*Hazard quociente, HQ*) (EUROPEAN COMMISSION, 2000; SANTOS *et al.*, 2020; SILVA, 2019).

No caso do risco à saúde humana considera-se a dose diária de fármaco tolerável, massa corporal e consumo de água diário, para encontrar o valor equivalente para a água de consumo (*Drinking Water Equivalent Level, DWEL*). Por fim, encontra-se a margem de exposição, medida do risco à saúde humana, pela razão entre o valor de DWEL e a concentração medida na água de consumo (WORLD HEALTH ORGANIZATION, 2012).

Dentre os efeitos observados de toxicidade, relata-se a exposição ao hormônio etinilestradiol associado à feminização de peixes (GROSS-SOROKIN; ROAST; BRIGHTY, 2006), o impedimento ao crescimento de células-tronco embrionárias por Carbamazepina e Atenolol, interferência na capacidade de reprodução de espécies aquáticas pelo Ibuprofeno, processo de fotossíntese limitado por antibióticos, bem como a promoção do desenvolvimento de genes resistentes a antibióticos, entre outros efeitos (PARIDA *et al.*, 2021; REICHERT *et al.*, 2019).

Pereira *et al.* (2020) identificaram alto risco por exposição à hormônios, antiepiléticos, anti-inflamatórios e antibióticos para algas, invertebrados e peixes em efluentes de ETE e águas superficiais. Para essas mesmas classes terapêuticas, Wilkinson *et al.* (2022) identificaram as maiores concentrações de ocorrência em nível global.

Particularmente, o grupo dos fármacos não pode ser considerado homogêneo como os outros grupos de CPEs, já que não necessariamente apresentam propriedades estruturais e físico-químicas semelhantes. Esses compostos possuem estrutura molecular complexa que pode variar em relação à massa molecular ou formato, bem como podem apresentar-se lipofílicos ou ter solubilidade em água de fraca à moderada, além de permanecer biologicamente ativos no meio ambiente (PATEL *et al.*, 2019).

Justifica-se a realização de estudos que identifiquem e quantifiquem concentrações de fármacos em matrizes aquáticas, assim como, investigações sobre os riscos associados à

exposição a esses contaminantes visando melhor elucidar os mecanismos, destino e efeitos desses contaminantes sobre o ecossistema aquático e a saúde humana.

No Quadro 1 mostra-se a classificação de alguns fármacos nas classes terapêuticas mais comumente investigadas quanto à ocorrência ambiental.

Quadro 1 – Classes terapêuticas de fármacos

Classe terapêutica	Fármacos	Identificador CAS	
Antibióticos	Amoxicilina,	26787-78-0	
	Azitromicina,	83905-01-5	
	Ciprofloxacino,	85721-33-1	
	Claritromicina,	81103-11-9	
	Norfloxacino	70458-96-7	
	Cefalexina	15686-71-2	
	Ampicilina	69-53-4	
Anti-inflamatórios ¹	Ofloxacino	82419-36-1	
	Diclofenaco,	15307-86-5	
	Ibuprofeno,	15687-27-1	
	Naproxeno,	22204-53-1	
Hormônios	Paracetamol	103-90-2	
	Estrona (E1),	53-16-7	
	17 β -estradiol (E2),	50-28-2	
Reguladores lipídicos	17 α -ethinylestradiol (EE2)	57-63-6	
	Genfibrozila,	25812-30-0	
	Simvastatina,	79902-63-9	
<i>Psicotrópicos</i>	Bezafibrato	41859-67-0	
	Antiepiléticos/ Antidepressivos	Carbamazepina	298-46-4
		Citalopram,	59729-33-8
		Fluoxetina,	54910-89-3
		Norfluoxetina,	83891-03-6
Sertralina		79617-96-2	
Ansiolíticos	Aprazolam,	28981-97-7	
	Lorazepam,	846-49-1	
	Zolpidem	82626-48-0	

Fonte: PEREIRA *et al.*, 2020; PUBCHEM, 2023; VALDEZ-CARRILLO *et al.*, 2020.

Nesse contexto, os resultados das pesquisas têm potencial de apontar os componentes prioritários para integrar regulamentações e normativas de controle e monitoramento, de modo a minimizar os riscos gerados à população e alcançar eficácia nas ações de preservação da qualidade do meio ambiente (GEISSEN *et al.*, 2015; PARIDA *et al.*, 2021).

3.2.1 Ocorrência de fármacos em matrizes aquáticas e aspectos limitantes do monitoramento

¹ Também identificados como Anti-inflamatórios Não Esteroides (AINES), são aplicados com função anti-inflamatória, analgésica e antitérmica (NHS, 2023).

Estima-se que cerca de 90% dos esgotos gerados em países em desenvolvimento sejam lançados deliberadamente sem qualquer tratamento nos corpos receptores (UNESCO, 2023). A carga lançada pode conter elevada carga orgânica, micro-organismos patogênicos, macronutrientes e CPEs comprometendo o padrão de qualidade dos corpos hídricos receptores e acarretando riscos de contaminação ao ecossistema aquático e seres humanos (GEISSEN *et al.*, 2015; KOSTICH; BATT; LAZORCHAK, 2014).

A promoção de ações de saneamento básico integrado é fundamental para minimizar os impactos negativos do lançamento de cargas poluidoras em águas superficiais. Desse modo, emerge uma preocupação sobre como gerenciar a ocorrência de CPEs em matrizes aquáticas com o intuito de potencializar a remoção desses compostos pelo tratamento aplicado em ETE e, assim, reduzir as concentrações lançadas nos corpos receptores e minimizar os impactos negativos atrelados a essa prática.

Concentrações de fármacos foram identificadas em águas superficiais e efluentes de tratamento de esgotos sanitários (CASTIGLIONI *et al.*, 2018; KOSTICH; BATT; LAZORCHAK, 2014; MONTAGNER *et al.*, 2019), água de abastecimento e consumo humano (REIS *et al.*, 2019; SANTOS *et al.*, 2020) em diferentes localidades.

Wilkinson *et al.* (2022) reportaram altas concentrações de fármacos em águas superficiais em regiões da África Subsaariana (2,94 – 5.660 ng. L⁻¹), América do Sul (2,4 – 7.580 ng. L⁻¹) e Ásia (4,3 – 4.080 ng. L⁻¹). Para a América do Norte e a Europa, ambos os continentes com recorrência de publicações de monitoramento, identificaram-se concentrações na faixa de 3,86 – 1.740 ng. L⁻¹ e 15,6 – 1.060 ng. L⁻¹, respectivamente. Nigéria, Tunísia e Costa Rica se destacaram entre os dez países com maior nível de concentração acumulada e pouquíssimos trabalhos de monitoramento (3, 2 e 1 trabalho, respectivamente).

Meléndez-Marmolejo *et al.* (2020) reportaram resultados de estudos de monitoramento de antibióticos na América Latina e constataram que dentre as concentrações reportadas, grande parte é proveniente de estudos no Brasil e México. Os autores identificaram que por vezes, as concentrações quantificadas em rios superavam as identificadas nos efluentes, indicando o lançamento de esgotos sem tratamento na água superficial.

Valdez-Carrillo *et al.* (2020) verificaram que apenas dez países da América Latina apresentam algum estudo sobre o monitoramento de fármacos em matrizes aquáticas e, ao todo, identificam 159 compostos. Além disso, apontam altas concentrações de Carbamazepina (830 µg. L⁻¹) e etinilestradiol (6,8 µg. L⁻¹) no Equador e Brasil.

Kot-Wasik *et al.* 2016 investigaram a ocorrência de 25 fármacos em ETE e ETA na Polônia em relação à sazonalidade, sendo a Carbamazepina identificada em todas as amostras

e considerada a mais persistente no meio ambiente, visto que sua concentração não variou consideravelmente conforme a estação do ano (3.508,6 ng. L⁻¹).

O estudo de Reis *et al.* (2019) identificou 18 fármacos em água superficial, destacando-se a concentração máxima 7.836 ng. L⁻¹ verificada para a Betametasona. Os autores também analisaram amostras de cinco ETA e identificaram 11 fármacos em água para consumo humano, com ocorrência de concentração máxima verificada no valor de 6.323 ngL⁻¹ para Prednisona. Frequentemente, as concentrações mínimas quantificadas estiveram abaixo do valor do limite de detecção do método. Os autores ainda alertam que os níveis identificados no estudo são superiores aos reportados em outras pesquisas.

Veras *et al.* (2019) investigaram a ocorrência dos anti-inflamatórios Diclofenaco e Paracetamol em água superficial em dois pontos de coleta, o primeiro em área preservada sem indícios de urbanização e o outro em área com intensa atividade antropogênica no estado de Pernambuco, Brasil. O Diclofenaco apresentou maior frequência de detecção, bem como maiores concentrações (concentração máxima de 0,193 mg. L⁻¹), enquanto o Paracetamol apresentou concentração máxima de 0,042 mg. L⁻¹. Os níveis verificados para ambos os fármacos foram mais altos no ponto urbanizado.

Oliveira et al. (2024) associaram altas concentrações de fármacos à intensa taxa de urbanização e lançamento de esgoto não tratado, em regiões do estado de São Paulo e Maranhão. Acetaminofeno foi verificado em concentrações de < 200 – 575,9 ng. L⁻¹, com alto risco ambiental. Naproxeno, identificado na faixa de <200 – 224,7 ng. L⁻¹ apresentou médio risco ambiental.

Fonseca et al. (2024) identificaram os anti-inflamatórios Naproxeno e Diclofenaco, nas concentrações máximas de 640 ng. L⁻¹ e 410 ng. L⁻¹, em águas superficiais de um rio localizado em região costeira do estado do Maranhão. Os autores reforçaram que a falta de saneamento influenciou os resultados verificados.

Rodrigues et al. (2024) investigaram a ocorrência de hormônios, fármacos, cafeína e PFAS no reservatório Três Marias, utilizado para cultivo de peixes, no estado de Minas Gerais. Altos níveis de PFAS foram identificados (9,40 – 15,2 µg L⁻¹), enquanto antibióticos e hormônios foram identificados em menores níveis. Destaca-se a ampla detecção de Ciprofloxacino, antibiótico vastamente associado a propagação da resistência antimicrobiana.

No Quadro 2 são apresentados os fármacos mais recorrentes em matrizes aquáticas, reportados na literatura.

Fármaco	Matriz aquática	Referência
Metformina, Atenolol, Ranitidina, Paracetamol, Cafeína, Ofloxacino, Trimetazidina, Ifosfamida, Ciclofosfamida, Carbamazepina, Eritromicina, Fluvoxamina, Ramipril, Nafronila, Levonorgestrel, Progesterona, Ácido Salicílico, Hidroclorotiazida, Furosemida, Cloranfenicol, Pravastatina, Losartana, Ketorprofeno, Diclofenaco, Ibuprofeno	Água superficial e tratada	Kot-Wasik et al. (2016)
Amoxicilina, Ampicilina, Cefalexina, Ciprofloxacino, Norfloxacino, Sulfametoxazol, Tetraciclina, Trimetoprima, Claritromicina, Clindamicina, Doxiciclina, Ofloxacina, Triclosan, Sulfametoxazol, Enrofloxacina, Azitromicina, Cloxacina, Eritromicina Oxacilina, Sulfadiazina, Sulfadimetoxina, Sulfametazina, Sulfanilamida Sulfatiazol, Oxitetraciclina, Cefaclor, Cefadroxilo	Água superficial e tratada	Meléndez-Marmolejo et al. (2020)
Amoxicilina, Cafeína, Cefalexina, Ciprofloxacino, Diclofenaco, Ibuprofeno, Norfloxacino, Paracetamol, Sulfametoxazol, Triclosan, Trimetoprima	Água superficial e para consumo humano	Montagner et al. (2019)
Betametasona, Claritromicina, Fenazona, Fenilbutazona, Fenofibrato, Fluconazol, Loratadina, Prednisona, Atorvastatina, Cafeína, Danofloxacino, Enoxacina, Enrofloxacina, Metformina, Genfibrozila, Ibuprofeno, Ketoprofeno e Norfloxacino	Água superficial e para consumo humano	Reis et al. (2019)
Atorvastatina, Betametasona, Cimetidina, Enrofloxacina, Fluconazol, Genfibrozila, Loratadina, Metformina, Norfloxacino, Omeprazol, Fenilbutazona, Prednisona	Água superficial e para consumo humano	Santos et al. (2020)
Diclofenaco, Paracetamol	Água superficial	Veras et al. (2019)
Acetaminofeno, Naproxeno, Diclofenaco	Água superficial	Oliveira et al. (2024)
Ácido Mefenâmico, Diclofenaco, Naproxeno	Água superficial	Fonseca et al. (2024)
Sulfametoxazol, Sulfadiazina, Trimethoprima, Ciprofloxacino, Cloranfenicol	Água superficial	Rodrigues et al. (2025)

Países com maior renda per capita, geralmente, apresentam melhor infraestrutura de saneamento, além de apresentarem restrições para a comercialização de fármacos, bem como o descarte, ou seja, mesmo com elevado consumo de fármacos, menores concentrações são verificadas no meio ambiente. Por outro lado, países de baixa e média renda per capita apresentam pior qualidade na prestação de serviços de saneamento, além da população ter acesso facilitado à aquisição de medicamentos e ausência de normativas e fiscalização sobre disposição de medicamentos vencidos, o que resulta em altas concentrações de fármacos em águas, especialmente, analgésicos e antibióticos (WILKINSON *et al.*, 2022).

Autores identificaram associação entre a concentração de ocorrência de fármacos com consumo de água, densidade populacional, persistência ambiental do composto, localidade investigada, sazonalidade, além do padrão de uso do fármaco (PARIDA *et al.*, 2021; PATEL *et al.*, 2019; SANTOS *et al.*, 2020).

Patel *et al.* (2019) citaram que o consumo de fármacos na Grécia foi afetado pela crise econômica entre os anos 2010 e 2014, sendo verificados picos de ocorrência de antipsicóticos, anti-hipertensivos e antidepressivos. Os autores também apontaram maiores cargas afluentes de antibióticos durante a estação de inverno, associadas ao aumento de consumo de fármacos para o tratamento de infecções respiratórias. O mesmo foi observado por Santos *et al.* (2020), em que os meses de inverno apresentaram alta concentração de fármacos devido ao baixo índice de chuvas e aumento do consumo de medicamentos para tratar doenças infecciosas. Os autores também identificaram que áreas com maior Índice de Desenvolvimento Humano (IDH) e Produto Interno Bruto (PIB) revelaram maior concentração de fármacos.

Apesar dos relatos apresentados, dados de monitoramento são escassos e, geralmente, provém de investigações realizadas em países da Europa como Alemanha, Reino Unido e Espanha, assim como, Estados Unidos e China (PARIDA *et al.*, 2021; WILKINSON *et al.*, 2022). Portanto, é primordial realizar o monitoramento dessas substâncias, especialmente, em localidades nas quais ainda não houve investigação sistemática, visto que dos 195 países reconhecidos pela ONU, aproximadamente, metade não reporta dados ou informações na literatura (OLDENKAMP *et al.*, 2022).

No Brasil, os estudos de monitoramento ocorrem próximos aos centros de pesquisa, concentrados em sua maioria em áreas mais populosas e desenvolvidas da região Sudeste. Dessa forma, regiões do interior do país comumente não estão incluídas em rotinas de monitoramento, apresentando estudos pontuais ou não reportando nenhum dado ou informação (REICHERT *et al.*, 2019).

Aspectos técnicos e financeiros impõem limitações à coleta e divulgação de dados. A definição dos métodos de amostragem varia com os estudos, dificultando a apresentação dos resultados de forma padronizada (KOSTICH; BATT; LAZORCHAK, 2014). Há uma dificuldade logística e financeira para coletar amostras em campo (geralmente em duplicata) de acordo com a técnica de amostragem definida (KOSTICH; BATT; LAZORCHAK, 2014) e, aponta-se, também, a dificuldade de interação com equipes de operação das ETE para acessar os dados de monitoramento da qualidade dos efluentes ou realizar a quantificação dos compostos, no caso de avaliações em esgotos sanitários (REICHERT *et al.*, 2019).

Sob o ponto de vista de análise e quantificação de compostos, há diferentes procedimentos analíticos disponíveis, o que torna difícil a interpretação de resultados. Além disso, é necessário considerar os custos de operação e manutenção dos equipamentos e aquisição de reagentes. Existem diversos compostos que ocorrem em baixíssimas concentrações e ainda não são quantificados, devido às limitações analíticas, o que faz com que

uma pequena parcela de fármacos seja mais frequentemente monitorada (OLDENKAMP *et al.*, 2022).

Por fim, o próprio comportamento dos fármacos devido as propriedades como persistência no ambiente, polaridade, capacidade de biodegradação e transformação, bem como interação em misturas e efeitos sinérgicos faz com que ainda sejam um desafio (GEISSEN *et al.*, 2015).

Poucos trabalhos investigaram a ocorrência de fármacos em esgotos sanitários, em geral, sendo a maioria das investigações voltadas para águas superficiais e águas de abastecimento público (GROSSELI, 2016; REICHERT *et al.*, 2019). No entanto, verifica-se uma tendência das concentrações máximas no seguinte sentido: água residual (esgoto bruto) > efluente de esgoto tratado > água superficial (VALDEZ-CARRILLO *et al.*, 2020). Isso reforça a necessidade de quantificar a ocorrência desses compostos em ETE, que são a principal fonte de contaminação ambiental.

3.2.2 Ocorrência de fármacos em ETE

Os fármacos são compostos reconhecidos por apresentarem estrutura estável, complexa, além de baixa volatilidade, hidrofobicidade e baixas concentrações de ocorrência (SANTOS *et al.*, 2020). Essas propriedades conferem resistência aos tratamentos biológicos para redução das concentrações nos esgotos sanitários, e justificam a ocorrência e persistência no meio aquático.

A ocorrência de fármacos juntamente com seus metabólitos é comum e, possivelmente, são gerados subprodutos da transformação dessas moléculas durante o tratamento dos esgotos sanitários nas ETE. Esses novos compostos podem apresentar potencial de risco ainda maior do que os compostos originários (SUBEDI *et al.*, 2017). O teor de fármacos verificado nos esgotos sanitários dependerá da quantidade consumida, bem como variáveis como taxa de excreção, produção de metabólitos e subprodutos e, no caso de tratamento, a capacidade de remoção da tecnologia empregada na ETE.

Um estudo realizado por Kostich *et al.* (2014) identificou diferentes classes terapêuticas em efluentes de ETE nos Estados Unidos. Os fármacos Ibuprofeno (anti-inflamatório) e os anti-hipertensivos Atenolol, Valsartan e Hidroclorotiazida apresentaram as maiores concentrações médias, com concentrações máximas quantificadas em 4.200 ng. L⁻¹, 3.000 ng. L⁻¹, 5.300 ng. L⁻¹ e 2.800 ng. L⁻¹, respectivamente.

Verlicchi *et al.* (2014) quantificaram a ocorrência de antibióticos e Carbamazepina no afluente e efluente de ETE na Itália. Os autores verificaram para Ciprofloxacino as concentrações médias de 2.200 ng. L⁻¹ e 630 ng. L⁻¹ correspondendo aos fluxos afluente e efluente, respectivamente. Norfloxacinó foi quantificado em 210 ng. L⁻¹ e 150 ng. L⁻¹ e Carbamazepina apresentou concentrações de 570 ng. L⁻¹ e 370 ng. L⁻¹.

Castiglioni *et al.* (2018) verificaram altas concentrações afluentes de antibióticos (1,6 µg. L⁻¹), anti-inflamatórios (5,9 µg. L⁻¹), fármacos cardíacos (2,8 µg. L⁻¹) e diuréticos (1 µg. L⁻¹) em ETE na Itália. Essas classes terapêuticas também apresentaram as maiores concentrações efluentes.

O estudo de Muriuki *et al.* (2020) investigou a ocorrência de antibióticos e antiretrovirais em três ETE no Quênia, que apresentaram faixa de concentração afluente de 0,72 – 1.463,50 µg. L⁻¹ e efluente de 0,93 – 847,05 µg. L⁻¹. Os valores mais altos de concentração foram verificados para o antibiótico Lamivudina e o antirretroviral Nevirapina apresentou as menores concentrações. No entanto, todos os fármacos avaliados apresentaram concentrações da ordem de µg. L⁻¹.

Subedi *et al.* (2017) identificaram 29 fármacos e produtos de cuidado pessoal em duas ETE na Índia e seis metabólitos, que demonstraram concentração de ocorrência sete vezes maior que a dos compostos iniciais. Os fármacos psiquiátricos Carbamazepina (concentração afluente média de 500 ng. L⁻¹) e Diazepam (concentração afluente média de 24 ng. L⁻¹), juntamente com o antiagregante plaquetar Clopidogrel (concentração afluente média 130 ng. L⁻¹) apresentaram alta resistência ao tratamento e houve formação de metabólito pelo Clopidogrel.

Vergili *et al.* (2019) quantificaram a ocorrência de fármacos no afluente e efluente de duas ETE, localizadas na Turquia e Alemanha. Verificaram-se concentrações afluentes de < 10 a 14.000 ng. L⁻¹, valor máximo identificado para o beta-bloqueador Sotalol. Destaca-se também os valores de 7.100 ng. L⁻¹ verificado para Carbamazepina e 6.600 ng. L⁻¹ para o Paracetamol. A faixa de concentração efluente variou de < 10 a 3.700 ng. L⁻¹, valor máximo da Carbamazepina. Os fármacos Sulfapiridina, Oxazepam e Diclofenaco apresentaram os maiores valores de coeficiente de risco (*Risk Quocient, HQ*), de 7.250, 310,5 e 215, respectivamente.

Montagner *et al.* (2018) desenvolveram um amplo monitoramento de CPEs em matrizes aquáticas na região de Campinas e São José do Rio Preto, região sudeste do Brasil. Os autores identificaram que o tratamento aplicado nas ETE investigadas não foi efetivo para a remoção da maioria dos compostos, especialmente, para o hormônio Estrona, o pesticida Atrazina e o produto industrial Bisfenol A. Uma avaliação de risco à saúde humana foi realizada para 18

compostos, dentre eles o Diclofenaco e o Paracetamol, e os resultados apontaram que apenas o hormônio 17β -estradiol poderia apresentar efeitos adversos.

No levantamento realizado por Rodrigues *et al.* (2025), fármacos pertencentes às classes terapêuticas dos anti-inflamatórios, antibióticos, psicotrópicos, cardiovasculares e, reguladores lipídicos, em correntes de águas urbanas, demonstraram aumento nas concentrações à jusante de ETE, atingindo valores de 185 ng. L^{-1} (reguladores lipídicos) a $26,218 \text{ ng. L}^{-1}$ (cardiovasculares).

Situação semelhante foi verificada por Kotke *et al.* (2024), cujo monitoramento em água superficial, na região de Coimbra, demonstrou aumento expressivo na concentração de fármacos à jusante do ponto de lançamento do efluente de ETE, com concentração total máxima de 3.500 ng. L^{-1} . A Carbamazepina apresentou alto coeficiente de risco ($HQ > 1$) à montante do ponto de lançamento, enquanto após, o valor aumentou para 53.

Netshithothole *et al.* (2024) detectaram concentrações alarmantes de fármacos no afluente de duas ETE, na África do Sul. Os valores foram consideravelmente altos para o antibiótico Sulfametoxazol ($9,11 - 77,33 \text{ } \mu\text{g. L}^{-1}$), Naproxeno ($8,21 - 73,11 \text{ } \mu\text{g. L}^{-1}$), Ibuprofeno ($3,55 - 59,84 \text{ } \mu\text{g. L}^{-1}$), Trimetoprima ($18,73 - 53,60 \text{ } \mu\text{g. L}^{-1}$), e Efavirenz ($4,08 - 10,09 \text{ } \mu\text{g. L}^{-1}$). Não foram detectadas concentrações de Naproxeno e Ibuprofeno nas amostras do efluente de uma das ETE, cuja remoção variou de 2% a 53% para os demais fármacos. Para a outra ETE, a eficiência de remoção variou de 12% a 85%, valores alcançados pelo Efavirenz e Trimetoprima, respectivamente.

Rodríguez-Rodríguez *et al.* (2024) identificaram que em estudos realizados na América Latina, os compostos mais comumente identificados em afluentes e efluentes de ETE são Ibuprofeno, Naproxeno, Acetaminofeno, Ketoprofeno, Gemfibrozil, Cafeína e Carbamazepina, que juntamente com Lovastatina, Trimetoprima, Norfloxacino e Risperidona, apresentaram alto coeficiente de risco, na faixa de 13,3 a 2.190 .

3.2.3 Remoção de fármacos em ETE

Ainda que diferentes fatores possam influenciar na ocorrência de fármacos, busca-se identificar as tecnologias de tratamento implantadas nas ETE que sejam eficazes para redução das concentrações desses compostos e descarte seguro no meio aquático. Pode-se afirmar que se trata de um desafio, visto que as ETE são projetadas para remoção de sólidos grosseiros, matéria orgânica e macronutrientes, além de microrganismos patogênicos. No entanto, moléculas mais complexas como os fármacos não são completamente removidas pelos

processos usualmente envolvidos no tratamento biológico (COUTO; LANGE; AMARAL, 2019).

As etapas do tratamento convencional realizado em ETE se subdividem em preliminar, primário, secundário e, terciário, utilizando uma associação de métodos biológicos e físico-químicos. O tratamento preliminar compreende os fenômenos físicos de gradeamento para remoção de sólidos grosseiros seguido pela remoção de areia por sedimentação. Em seguida, inicia-se a etapa primária com a finalidade de separar a matéria orgânica particulada presente nos esgotos sanitários originando o lodo primário por meio da sedimentação. Algumas ETE, aplicam o processo de coagulação nessa etapa do tratamento (GROSSELI, 2016; METCALF & EDDY, 2014).

A etapa de tratamento secundário consiste, na maioria dos casos, na aplicação de tratamento biológico por meio de rotas anaeróbias e/ou aeróbias, cujo principal objetivo é a remoção de matéria orgânica solúvel por meio da atividade microbiológica. Além disso, quando associado a mecanismos específicos terciários, as ETE podem promover a remoção de macronutrientes (fósforo e nitrogênio), geralmente, por meio da associação de processos biológicos e físico-químicos (COUTO; LANGE; AMARAL, 2019). A eficiência dos processos se relaciona com as características dos compostos presentes na água residuária frente aos parâmetros operacionais como tempo de retenção de sólidos (TRS), TRH (tempo de retenção hidráulica), temperatura, pH, coeficiente de partição (K_{ow}) e arranjo do tratamento aplicado (METCALF & EDDY, 2014).

A remoção biológica da matéria orgânica solúvel (etapa secundária do tratamento) é alcançada em reatores biológicos como biodigestores anaeróbios, reatores anaeróbios de fluxo ascendente e manta de lodo (UASB), lagoas de estabilização, sistemas de lodos ativados, filtros biológicos, *wetlands* construídos, dentre outros. Diferentes combinações entre as tecnologias podem ser configuradas a depender do objetivo do tratamento. Para a remoção de micro contaminantes inorgânicos, como metais pesados, e orgânicos específicos, como os fármacos, a aplicação somente desses processos não garante eficiências elevadas. Entretanto, nas pesquisas, a avaliação da remoção dos fármacos é usualmente reportada nas etapas de tratamento secundário (GONZÁLEZ PEÑA; LÓPEZ ZAVALA; CABRAL RUELAS, 2021).

O tratamento terciário, avançado, ocorre para complementar o tratamento secundário para a remoção de compostos específicos, mais resistentes e persistentes à remoção biológica. Essa etapa pode compreender o emprego de tecnologias físico-químicas como os processos oxidativos avançados (UV/H₂O₂), Fenton (Fe⁺² + H₂O₂) ou foto-Fenton (UV + Fe⁺² + H₂O₂), eletrocoagulação ou filtração com membranas (FOSSO-KANKEU, 2019; LOPES *et al.*, 2016).

Verlicchi *et al.* (2012) verificaram a eficiência de remoção de 118 fármacos em esgotos sanitários brutos e efluentes de tratamento secundário de um sistema de lodo ativado e biorreator com membranas. Os autores identificaram maiores concentrações de anti-hipertensivos, beta bloqueadores e analgésicos/anti-inflamatórios no efluente secundário e em geral, o biorreator com membranas apresentou maior eficiência de remoção dos fármacos.

Li *et al.* (2019) investigaram a degradação dos anti-inflamatórios: acetaminofeno (paracetamol), ibuprofeno, naproxeno, bezafibrato, antipirina e diclofenaco em amostras provenientes de um processo de lodo ativado. Verificou-se que naproxeno, acetaminofeno e ibuprofeno foram completamente removidos, enquanto outros compostos como diclofenaco tiveram aumento de concentração. Os autores destacaram que a adição de agentes flocculantes e o tratamento de desinfecção com aplicação de radiação UV são indispensáveis para a remoção dos contaminantes. As concentrações afluentes variaram de 0 a 510,8 ng. L⁻¹ e as efluentes variaram de 0,1 a 55 ng. L⁻¹.

Chen *et al.* (2021) identificaram que a combinação dos tratamentos terciários com aplicação de ozônio, filtração com membrana cerâmica e adsorção em leito de carvão ativado foi capaz de promover eficiências de remoção de 98,5%, 95,4% e 91,1% para efluente secundário contendo concentrações de fármacos e produtos de cuidados pessoais, disruptores endócrinos e compostos de odor, respectivamente.

Grosseli (2016) investigou a ocorrência e capacidade de remoção de 12 fármacos das classes terapêuticas anti-inflamatório, psicotrópico, beta-bloqueadores em duas ETE localizadas no estado de São Paulo. A ETE localizada na cidade de Jundiaí emprega tratamento aeróbio por lagoas aeradas de mistura completa seguido por lagoa de decantação e demonstrou eficiência de remoção maior que 70% para cinco compostos, sendo identificada remoção inferior a 20% para Carbamazepina. A ETE localizada na cidade de São Carlos constituída por um reator UASB seguido de unidade de flotação, apresentou altas eficiências de remoção apenas para a Cafeína. Para o Ibuprofeno e a Carbamazepina as remoções foram inferiores a 20%. Portanto, o tratamento aeróbio se mostrou mais eficaz na remoção dos fármacos. Adicionalmente, o autor identificou que no processo de tratamento foram gerados metabólitos de Diclofenaco, Carbamazepina, Propanolol, Atenolol e Ibuprofeno.

Subedi *et al.* (2017) verificaram a ocorrência de 29 fármacos e seis metabólitos no afluente de duas ETE, na Índia. Uma ETE emprega o processo de lodo ativado e o arranjo tecnológico da outra é uma combinação de reatores UASB e processo de lodo ativado, seguido por clarificação e desinfecção. Carbamazepina, Diazepam, Clopidogrel se mostraram resistentes ao tratamento por lodo ativado. Por outro lado, a combinação de tratamento

anaeróbio e aeróbio (UASB + lodo ativado), seguida por pós-tratamento, proporcionou uma remoção de até 95% desses compostos.

Valdez-Carrillo *et al.* (2020) constataram a concentração de 830 $\mu\text{g. L}^{-1}$ de Carbamazepina em água superficial no Equador, um valor superior ao verificado em outros países. Os autores identificaram recorrência na frequência de detecção desse fármaco em efluente de ETE ou águas superficiais e a partir das estimativas, esse composto apresentou o maior risco ecotoxicológico dentre todos os fármacos avaliados.

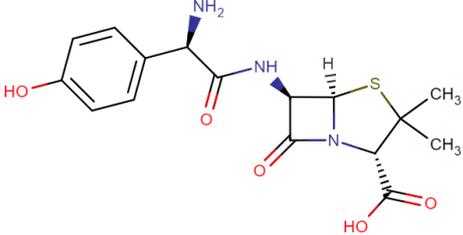
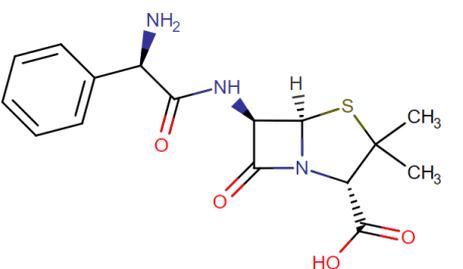
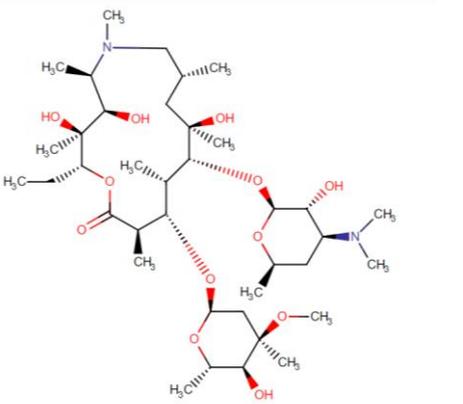
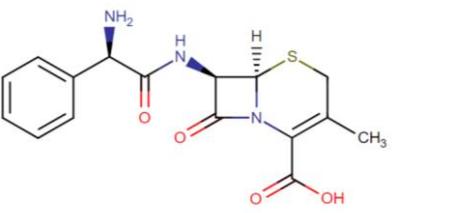
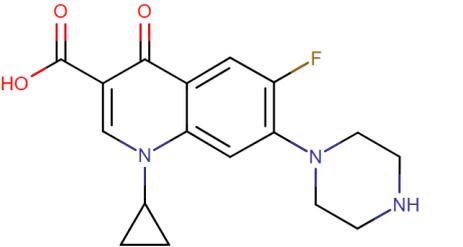
Pivetta *et al.* (2020) verificaram que a concentração de Carbamazepina na corrente afluyente ou efluente de ETE não apresentou variação significativa com o tratamento empregado, independente das tecnologias aplicadas. As concentrações identificadas não apontaram risco ecotoxicológico, embora tenha sido detectada concentração máxima efluente de 3.000 ng. L^{-1} . Os autores destacam que caso fosse considerado o descarte sem tratamento e sem o fator de diluição, as concentrações poderiam ocasionar riscos ao ecossistema.

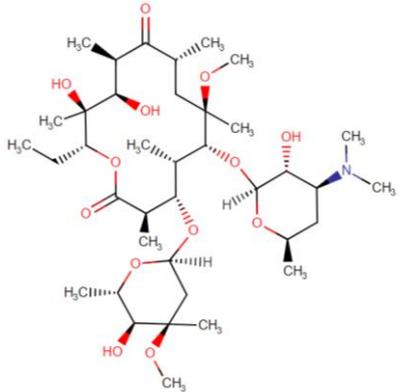
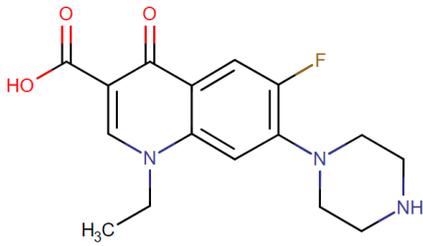
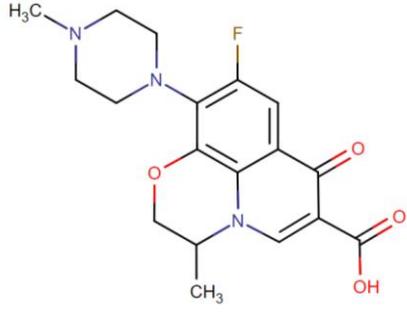
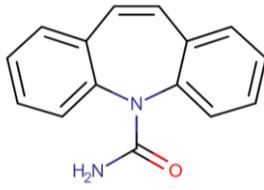
Verlicchi *et al.* (2012) identificaram que as drogas psiquiátricas Carbamazepina, Fluoxetina e Diazepam apresentaram alto risco ambiental, juntamente com as classes de anti-inflamatórios e antibióticos, sendo esses os fármacos Eritromicina, Ofloxacino, Sulfametoxazol, Claritromicina, Amoxicilina, Tetraciclina e Azitromicina.

Quanto à ocorrência de antibióticos, Reichert *et al.* (2019) verificaram que na América Latina, a resistência a antibióticos foi reportada especialmente no Brasil e sugerem uma lista prioritária para o monitoramento de 13 fármacos, dentre eles a Amoxicilina, Norfloxacino e Ampicilina.

Na Tabela 1 apresenta-se propriedades dos fármacos Ampicilina (AMP), Amoxicilina (AMOX), Azitromicina (AZT), Cefalexina (CEF), Ciprofloxacino (CIP), Claritromicina (CLA), Norfloxacino (NOR), Ofloxacino (OFL), juntamente com o psicoativo Carbamazepina (CMZ).

Tabela 1 – Propriedades dos fármacos selecionados para o estudo

Fármaco	Classe de Antibiótico	Log Kow ³	Solubilidade em água (25 °C) ² (mgL ⁻¹)	Estrutura ¹	Massa Molecular (g mol ⁻¹) ¹
AMOX	b-Lactâmico	0,97	3430		365.404
AMP	b-Lactâmico	1,35	10100		349.405
AZT	Macrolídeo	4,02 ²	2,37		748.9845
CEF	b-Lactâmico	0,65	1789 ³		347.389
CIP	Fluroquinona	0,28	11480		331,3415

Fármaco	Classe Antibiótico	de	Log Kow ³	Solubilidade em água (25 °C) ² (mgL ⁻¹)	Estrutura ¹	Massa Molecular (g mol ⁻¹) ¹
CLA	Macrolídeo	3,16 ²	1,693			747,9534
NOR	Fluroquinona	-1,03	177900 ³			319,3308
OFL	Fluoroquinona	-2,00	6762000 ³			361,3675
CMZ	Antiepilético	2,45	17,66 ³			236,2686

¹DrugBank (2023);²Pubchem (2023);³ChemSpider(2023).

Taoufik *et al.* (2020) apontaram que o principal tipo de tratamento empregado para a remoção de fármacos reportado na literatura científica é o físico-químico. Nesse sentido, a categoria dos Processos Oxidativos Avançados (POA), foto catálise, oxidação química com ozônio, aplicação de radiação UV tem se mostrado efetivas, mas, sua ampla aplicação em ETE é limitada pelo alto custo energético e operacional. Menores consumos energéticos e menores custos e complexidade operacionais podem ser obtidas pelos tratamentos biológicos, muito embora, maior área seja requisitada sem garantia de elevadas eficiências.

Na Figura 2, indica-se as principais tecnologias de remoção aplicadas em ETE. O termo biodegradação se refere à degradação dos compostos por meio da ação de diferentes microrganismos.

Figura 2 – Comparação dos tratamentos aplicados para a remoção de fármacos

Adsorção	Floculação	Lodo ativado
Microfiltração/ Nanofiltração	Oxidação	Biodegradação
Osmose Reversa	Ox. Eletroquímica	Reator de Membrana
	Ozônio	<i>Wetlands</i>
	Fenton	
Físico	Químico	Biológico
Simples configuração Custo benefício Efluente de alta qualidade Remoção de moléculas menores	Simplicidade e custo benefício Remoção de contaminantes insolúveis Mineralização de CPEs Desintoxicação Geração de produtos biodegradáveis	Simplicidade e custo benefício Ambientalmente viável Minimização de reações paralelas na remoção de CPEs Eliminação de patógenos Alta eficiência de degradação de compostos

Fonte: Adaptado de TAOUFIK *et al.* (2020).

Identifica-se que tipicamente, para garantir uma eficiência de remoção adequada é necessário associar o tratamento secundário a uma ou mais tecnologias de tratamento terciário. No entanto, os custos para a aquisição, operação e manutenção de tais aparatos é um fator restritivo ao alcance de um efluente devidamente tratado, com mínimas concentrações de fármacos.

3.2.4 Marcos Regulatórios

O conhecimento de aspectos legais possibilita compreender como as agências e órgãos ambientais tem abordado a questão da ocorrência de CPEs. A conjuntura regulatória atual considera os CPEs de forma geral e parte de evidências sobre os riscos calculados quando ocorrem nos compartimentos ambientais. Para as matrizes aquáticas, podem ser considerados riscos ecotoxicológicos para ecossistemas aquáticos ou riscos à saúde humana, a partir da consideração da ingestão da água contaminada.

Destaca-se a lista prioritária desenvolvida pela agência *Global Water Research Coalition*, que a partir de inúmeras publicações prévias sobre a ocorrência de fármacos em matrizes aquáticas identificou critérios para rotular os compostos em três categorias de prioridade para investigações: Classe 1 – prioridade alta, Classe 2- prioridade média, Classe 3 – prioridade baixa. Assim, a lista foi considerada por pesquisadores como critério de seleção de compostos para investigações quanto a aplicação de métodos analíticos, ocorrência, remoção e análises de potenciais riscos da ocorrência (DE VOOGT *et al.*, 2009).

No ano 2000, a “*Directive 2000/60/EC*” foi proposta com o objetivo de estabelecer práticas para a melhoria da qualidade ecológica de águas superficiais e subterrâneas por meio da redução de emissões de compostos perigosos nessas matrizes (EUROPEAN COMMISSION, 2000). As substâncias prioritárias presentes na lista de monitoramento incluíram CPEs das classes de herbicidas, hidrocarbonetos policíclicos aromáticos (HPAs), substâncias Per e Polifluoroalquil (PFAS), metais tóxicos e outros.

Na atualização do ano de 2009, houve a inclusão de 28 novos componentes, como o interferente endócrino Bisfenol A, agentes quelantes, novos herbicidas e também os fármacos Carbamazepina e Diclofenaco (EUROPEAN COMMISSION, 2009). Destaca-se que no ano de 2013 uma nova atualização contou com a inclusão dos fármacos Ibuprofeno e Ciprofloxacino (PARIDA *et al.*, 2013) e em 2015 foram incluídos os hormônios 17-alfa-etinilestradiol (EE2), 17-beta-estradiol (E2) e Estrona (E1), juntamente com os antibióticos macrólidos Eritromicina, Azitromicina e Claritromicina (EUROPEAN COMMISSION, 2015).

No ano de 2018, o Diclofenaco foi retirado da lista e incluiu-se a Amoxicilina (EUROPEAN COMMISSION, 2018). Esse fármaco, assim como o Ciprofloxacino foi mantido na lista atualizada no ano 2020, na qual incluíram-se os fármacos Sufametoxazol, Trimetoprima, Venlafaxina, Clotrimazol, Fluconazol, Miconazol, além de novos pesticidas e fungicidas (EUROPEAN COMMISSION, 2020).

A Agência de Proteção Ambiental dos Estados Unidos (USEPA) elabora periodicamente a *Contaminants Candidate List (CCL)*, na qual publica substâncias identificadas em matrizes aquáticas, mas, que ainda não fazem parte das normativas e

regulamentações de qualidade de água para consumo expedidas pela agência. Nos anos de 2016 e 2021, foram publicadas as listas CCL-4 e CCL-5, que incluem pesticidas, toxinas de origem microbiológicas, produtos de desinfecção e os fármacos: Eritromicina, Diazenon e Fluconazol. No ano de 2023, a agência consentiu ao público geral a indicação de substâncias candidatas à inclusão na sexta lista de prioridade, que está em fase de elaboração (EPA, 2023).

A Organização Mundial de Saúde (OMS) propôs o documento *Pharmaceuticals in Drinking Water* (WORLD HEALTH ORGANIZATION, 2012) para reportar concentrações de ocorrência de fármacos, discutir o avanço nas metodologias analíticas e avaliar potenciais riscos à saúde humana devido à exposição em água de consumo. Naquela época, as concentrações reportadas não ofereceram riscos, visto que raramente haviam sido identificados fármacos nas matrizes avaliadas e, quando presentes, apresentavam níveis baixos ($< 50 \text{ ng. L}^{-1}$).

Posteriormente, a OMS alertou para a ocorrência de pesticidas e reforçou que as maiores concentrações de fármacos ocorrem em efluentes de ETE ou em águas superficiais contaminadas pelo lançamento de efluentes industriais indevidamente tratados. A falta de padronização nos protocolos de amostragem e análise, o alto custo e disponibilidade limitada de tecnologias de detecção foram identificados como os principais limitantes para a implementação de programas de monitoramento (WORLD HEALTH ORGANIZATION, 2022).

Outras regiões têm mostrado iniciativas com base nas regulamentações já citadas. Parida *et al.* (2021) cita que o uso de Triclosan foi limitado na Argentina, Chile e pela Comunidade Andina. Na Austrália, foi proposta uma regulação para ocorrência de CPÉs em esgotos tratados que inclui Ciprofloxacino, Codeína, Carbamazepina, entre outros (NRMMC, 2008).

No Brasil, não há regulamentações específicas para o monitoramento e limitação da concentração de fármacos nas matrizes ambientais, ainda que as publicações sobre o tema tenham crescido nos últimos anos. No entanto, esparsas iniciativas foram iniciadas em relação aos pesticidas. A Portaria MS nº 2.914/2011 (BRASIL, 2011a) que dispõe do padrão de potabilidade da água para consumo humano, considerava apenas 23 pesticidas, mesmo diante de propostas de inclusões pela CETESB (ROQUETTI, 2010; MONTAGNER; VIDAL; ACAYABA, 2017). No ano de 2021, a Portaria nº 888/2021 (BRASIL, 2021) alterou a anterior, passando a incluir valores máximos permitidos para 40 pesticidas e metabólitos.

Sobre o lançamento de efluentes contaminados, ainda não há limites regulatórios para fármacos que as operadoras de ETE devam monitorar e obedecer. A Resolução CONAMA nº 430/11 (BRASIL, 2011b) estipula os valores máximos permissíveis para o lançamento de esgotos sanitários tratados em corpos d'água, e regulamenta a qualidade por meio do

monitoramento dos parâmetros: pH, temperatura, concentração de materiais sedimentáveis, Demanda Bioquímica de Oxigênio (DBO), teor de óleos e graxas, além de alguns metais, íons e substâncias orgânicas, porém, não há valores regulamentados para fármacos.

Adicionalmente, verifica-se que a diversas normativas existentes avaliam a ocorrência dos compostos frente ao risco que podem ofertar quando presentes em água para consumo humano, e nem sempre consideram os riscos ecotoxicológicos ofertados à biota aquática. Uma vez que não há um protocolo único e taxativo que pressione diferentes setores a trabalhar rumo à redução das emissões, o mercado farmacêutico permanece sem considerar o impacto ambiental do lançamento dos fármacos pós consumo como fator restritivo à comercialização e soma-se a isso o fato de que poucos estudos são realizados em prol do desenvolvimento de fármacos ambientalmente sustentáveis (GONZÁLEZ PEÑA; LÓPEZ ZAVALA; CABRAL RUELAS, 2021).

3.3 O USO DE MODELAGEM EM SISTEMAS DE REMOÇÃO DE FÁRMACOS

Estações de Tratamento de Esgotos são sistemas altamente complexos e desafiadores para o gerenciamento, dado que as etapas de remoção dos poluentes, nutrientes e matéria orgânica, envolvem uma série de reações físico-químicas e bioquímicas, interações entre microrganismos e contaminantes e, comportamento dinâmico da qualidade do efluente em relação a diferentes eventos (HAMADA; ZAQOOT; SETHAR, 2024; LI et al., 2024). Portanto, esses empecilhos devem ser mitigados para a garantia da qualidade do efluente final.

O desafio se torna ainda mais notório para a remoção de poluentes emergentes, como os fármacos. Isso se deve à alta quantidade de compostos existentes e a heterogeneidade em suas propriedades estruturais, moleculares e físico-químicas, que os torna resistentes aos tratamentos de remoção. Frente à complexidade do sistema de tratamento e o conhecimento limitado sobre os mecanismos de degradação, interação com substâncias e formação de metabólitos desses poluentes e pouca viabilidade de análises experimentais, a modelagem matemática passa a ser uma estratégia fundamental para complementar o monitoramento e potencializar a eficiência dos tratamentos realizados.

Nesse sentido, modelos fenomenológicos contemplam uma abordagem interessante para a predição da qualidade do efluente, remoção de nutrientes e contaminantes. Esses modelos se baseiam no conhecimento prévio dos fenômenos, considerando balanços de massa e dinâmicas de reação em processos de transporte, biotransformações e sorção, dependendo de um processo extenso de calibração baseado em experimentos de bancada.

Os modelos do tipo ASM (*Activated Sludge Model*) são os mais reconhecidos na academia e na indústria pelo pioneirismo na representação dos processos de tratamento de efluentes, com foco na representação da degradação da matéria orgânica e remoção de macronutrientes, como o Nitrogênio, por processo biológico. Ao longo dos anos, diversas expansões foram propostas para acrescentar outros mecanismos relevantes como remoção de fósforo, nitrificação-desnitrificação, entre outros, confirmando que esses modelos proporcionaram um alto ganho de conhecimento acerca dos mecanismos e dinâmicas do processo (DUARTE et al., 2024; JIANG et al., 2024; MENDES et al., 2023).

Em relação à ocorrência de fármacos, Snip et al. (2014) propuseram a integração da expansão ASM-X (*Activated Sludge Model for Xenobiotic*) na plataforma BSM (*Benchmark Simulation Model*), com o objetivo de modelar a carga de fármacos a partir de princípios fenomenológicos. Uma extensa quantidade de processos e taxas de biotransformação possibilitaram investigar o impacto de diferentes estratégias operacionais nos processos de sorção, dessorção em condições anaeróbias e anóxicas. No entanto, a capacidade de generalização do modelo é questionável, visto que a inclusão de novos fármacos exigiria a atualização das equações de taxas de reação, inclusão de novos processos de remoção e, ainda, a aplicação em uma nova ETE exigiria a correção de parâmetros relacionados ao transporte dos fármacos.

Embora esses modelos sejam baseados em conhecimento científico e prático, fazem simplificações dos processos e trazem considerações de idealidade que nem sempre refletem as condições reais dos sistemas. Além disso, o fato de serem implementados para casos específicos e o extenso número de parâmetros cinéticos e estequiométricos para calibração restringe a adaptação desses modelos a outras condições. Assim, nos últimos dez anos tem sido observado um declínio de publicações baseadas em modelos fenomenológicos, ao passo que observa-se um aumento no uso de Aprendizado de Máquina (DUARTE et al., 2024).

Outra questão relevante é o fato de que a complexidade do modelo aumenta conforme a quantidade de espécies a ser investigada, o que torna necessário utilizar modelos de simulação molecular e mecânica quântica, cujas variáveis serão propriedades quânticas e descritores moleculares como ângulo de ligação, distância da ligação, ângulo de torção, ω (índice de eletrofilicidade), Energia de ativação, constante de Planck e constante de Boltzman, para citar algumas. Por esses modelos é possível calcular as posições dos átomos, distinguir isômeros e calcular a reatividade química, características relevantes que influenciam a remoção dos compostos (HE et al., 2021).

Métodos dessa abordagem — DFC (*Density functional theory*) e QSAR (*Quantitative structure-activity relationship*) — demonstraram eficiência na identificação de variáveis mais correlacionadas com a degradação de fármacos e produtos de cuidados pessoais por processo de lodo ativado; bem como na predição da oxidação de compostos orgânicos pelo radical sulfato (HE et al., 2021). No entanto, esses modelos demandam alto custo computacional e são altamente complexos (ACOSTA-ANGULO et al., 2024; LI et al., 2024).

Diante desse cenário, os modelos de Aprendizado de Máquina têm mostrado alto potencial para otimizar as condições operacionais dos sistemas de tratamento de efluentes e prever níveis de poluentes, provando a capacidade de lidar com diferentes tipos de dados, aprender funções complexas e integrar diferentes variáveis. Assim, no contexto de aumento de poder computacional associado a diferentes formas de aquisição de dados, esses métodos propiciam objetividade e velocidade para identificar as relações entre entradas e saídas e fornecer modelos adaptáveis (TAOUFIK et al., 2022).

Desse modo, a próxima sessão se propõe a introduzir conceitos fundamentais para a compreensão do campo de Aprendizado de Máquina e dos modelos baseados em árvores empregados nesta pesquisa. Esse foi um modo para fortalecer, nesta tese, a aproximação entre as áreas de modelagem e meio ambiente, cuja falta de integração já foi apontada previamente (FAN et al., 2018; HADJIMICHAEL; COMAS; COROMINAS, 2016; TAOUFIK et al., 2022).

3.4 APRENDIZADO DE MÁQUINA

Aprendizado de Máquina, do inglês *Machine Learning*, é a área da Inteligência Artificial que propõe técnicas/modelos de aprendizagem para produzir inferências, associações complexas e predições a partir da identificação de padrões nos dados. Por meio da combinação de áreas da computação, estatística e otimização são propostos algoritmos com capacidade de conceber regras de aprendizado a partir de exemplos prévios e se adaptar a novos padrões (MICHIE *et al.*, 1995).

Tipicamente, as técnicas são identificadas para problemas de aprendizado não supervisionado ou aprendizado supervisionado. O aprendizado supervisionado se aplica a casos cujo objetivo é prever o valor de uma resposta a partir de medições de entrada, sendo que para cada observação das variáveis de entrada há uma resposta associada para classificação ou predição. Já o aprendizado não supervisionado ocorre quando não se conhece a resposta associada às variáveis de entrada. A predição é realizada unicamente pela descoberta de

associações e padrões existentes entre as variáveis de entrada (HASTIE; FRIEDMAN; TIBSHIRANI, 2001).

Taoufik *et al.*(2022) apresentam oito etapas para a concepção de um projeto de Aprendizagem de Máquina: identificação do problema, coleta de dados, exploração de dados, preparação de dados, escolha do modelo, avaliação do modelo e implantação do modelo. A preparação dos dados, ou pré-tratamento, tem a finalidade de filtrar variáveis ou observações, transformar a escala dos dados, padronizar nomes e identificar valores faltantes para decidir como geri-los. Com isso, colunas de variáveis ou linhas de observações são retiradas ou ajustadas na base, para preservar a qualidade e relevância dos dados e facilitar o reconhecimento de padrões pelo algoritmo (BISHOP; NASRABADI, 2006).

A escolha do modelo parte da correta identificação do problema. Uma vez definida a abordagem como aprendizado supervisionado, por exemplo, a escolha deverá ser feita entre as técnicas dessa categoria. Nesse caso, podem ser citadas técnicas de Regressão linear, Árvore de Decisão, Floresta Aleatória, Máquinas de Vetores de Suporte, entre outras que permitem obter uma resposta final numérica ou categórica.

No caso de aprendizado não supervisionado, aplicam-se técnicas para redução de dimensionalidade de dados como Análise Linear Discriminante ou Análise de Componentes Principais, além de técnicas de agrupamento não hierárquico como *k-means* ou agrupamento hierárquico como Dendrograma. Essas técnicas também podem ser utilizadas de forma exploratória para investigar as relações e tendências entre variáveis. Assim, ao conhecer o tipo de técnica adequada, implementa-se o algoritmo que automatiza a construção do modelo.

Na etapa de treinamento do modelo, uma parcela dos dados de entrada é inserida no algoritmo que irá identificar a tendência dos dados, realizando o processo de aprendizado. Nesse caso, a parcela de dados utilizada para ajustar o modelo chama-se de conjunto de treinamento. O resultado final de predição ou classificação obtido no treinamento deve ser avaliado com base em alguma métrica escolhida.

Posteriormente, realiza-se avaliação de desempenho do modelo treinado quando executado para novos dados. Assim, uma parcela menor dos dados (conjunto de teste) é utilizada para testar o ajuste a fim de avaliar a capacidade de generalização do modelo. Ou seja, simular a entrada de novos dados no modelo e quando necessário, devem ser feitas alterações para garantir que o modelo desempenhe adequadamente nos dados de teste. Uma vez identificado que o modelo aprendeu e apresenta desempenho satisfatório para novos dados, poderá ser colocado em operação.

Os algoritmos de Aprendizado de Máquina exercem um papel importante no fomento do uso de tecnologia na área ambiental, pois possibilitam prever a remoção de diferentes tipos de poluentes e aumentar a eficiência dos processos sem, necessariamente, elevar os custos financeiros (FAN *et al.*, 2018). Esse aspecto se justifica porque as técnicas podem desobrigar a necessidade de uma rotina frequente de análises experimentais e, conseqüentemente, dispêndios para aquisição de reagentes e equipamentos laboratoriais. No entanto, aponta-se que ainda há necessidade de mais estudos nesse sentido já que em geral os dados aplicados para a construção dos modelos partem de escala laboratorial, e não da investigação em plantas reais de tratamento (TAOUFIK *et al.*, 2022).

A aplicação de modelos de aprendizado no contexto de saneamento foi investigada previamente pela autora dessa tese quanto ao potencial enquanto ferramenta de planejamento, controle e gerenciamento de Resíduos Sólidos Municipais (RSM) na região da América Latina e Caribe. Positivamente, alguns países da região apresentam bases de dados abertas em relação ao monitoramento de RSM. No entanto, a qualidade dos dados é um fator crítico, visto que não há um processo seguro de gestão das bases. Portanto, os dados devem ser tratados e avaliados para reduzir as incertezas e, de fato, serem eficientes na proposição de modelos preditivos do padrão de geração de resíduos.

Hadjimichael *et al.* (2016) apontam que os métodos de Aprendizado de Máquina possibilitam criar uma interface para transformar dados em conhecimento para a tomada de decisão no setor de águas urbanas. Porém, uma falta de integração entre a área ambiental e computacional foi observada. Os estudos investigados lidam com problemas complexos e, apenas cerca de 5% consideram os resultados obtidos com base em questões práticas de planejamento e decisão no setor.

3.3.1 Métodos de Aprendizado de Máquina no controle e monitoramento de contaminantes emergentes

As limitações para obtenção de dados de monitoramento de poluentes no ambiente, assim como dados e informações sobre as eficiências de remoção dessas substâncias durante o tratamento de esgotos sanitários tornam necessário utilizar meios alternativos para avançar na compreensão do destino desses compostos no meio ambiente. A modelagem tem sido uma alternativa que pode auxiliar a superar o problema da falta de dados na área sem a necessidade de obter frequentemente novos dados de ocorrência de fármacos (VERLICCHI *et al.*, 2014).

Nesse contexto, a escassez e má distribuição de dados da ocorrência dos contaminantes é um desafio amplamente discutido na literatura científica, cuja solução encontra-se em duas possibilidades: realização de extensas pesquisas de monitoramento “*in loco*”, ou, estimativas das concentrações por meio de técnicas de modelagem a partir de dados prévios. O uso de modelos é essencialmente necessário no cenário para o qual, além da escassez de dados, pouco se conhece das relações entre os fatores que influenciam a ocorrência dos contaminantes e qual a expectativa para um futuro de mudanças globais (ACUÑA et al., 2020).

Assim, os modelos possibilitam investigar de modo assertivo relações entre diferentes variáveis que possam afetar os níveis de contaminantes nas matrizes ambientais, além de viabilizar a investigação de diferentes cenários de contaminação, sendo uma ferramenta de apoio para a definição de estratégias preventivas e remediadoras dos impactos negativos a médio e longo prazo.

Os modelos de aprendizado possibilitam ainda, expandir a forma como o problema é descrito, seja pela inclusão de variáveis até então não consideradas por modelos convencionais, ou mesmo, capacidade de lidar com situações contrastantes como ausência de dados, situação mais comum, ou grande quantidade de dados (WU; LI; XIONG, 2021). Desse modo, esses modelos proporcionam resultados assertivos a partir de menos recursos, ampliando as possibilidades de acompanhamento dos compostos no ambiente.

Portanto, o uso de modelos preditivos constitui uma alternativa confiável para expandir a avaliação da ocorrência de compostos indesejados nas matrizes ambientais. Até o momento, modelos preditivos aplicados para relacionar a concentração de ocorrência de fármacos no efluente do tratamento de esgoto a partir de vendas de medicamentos consideram a mesma relação matemática entre as variáveis e a eficiência de remoção das ETE (VERLICCHI *et al.*, 2014).

Ao mesmo tempo, algoritmos de Aprendizado de Máquina foram aplicados recentemente com sucesso na predição da remoção de poluentes de matrizes aquáticas a partir das variáveis do processo de remoção. Wu *et al.* (2021) investigaram a variação temporal nas concentrações de *Polychlorinated biphenyl* (PCBs), pesticidas e fármacos, dentre outros contaminantes, na região dos Grandes Lagos (EUA), a partir de um modelo de Rede de memória de curto prazo (*Long Short Term Memory*, *LSTM*), identificaram tendências sazonais, em especial um pico de concentração de PCB's no mês de julho.

Taoufik *et al.* (2022) apontaram a vasta aplicação de modelos de Redes Neurais Artificiais (ANN), Florestas Aleatórias, Árvores de Decisão e Máquinas de Vetores de Suporte (SVM) para a predição da remoção de fármacos e outros micropoluentes pelas técnicas

Processos Oxidativos Avançados (POA) e Adsorção. A partir de variáveis como pH, dosagem de componente, concentração de componente, tempo de reação e outras, modelos obtidos com R^2 acima de 0.90 reforçam a capacidade preditiva e o potencial de extração de padrões para melhorar o desempenho do tratamento de remoção através de técnicas de aprendizado de máquina (DOLATABADI *et al.*, 2020; FELICIANO *et al.*, 2020; SMAALI *et al.*, 2021; VAKILI *et al.*, 2019).

Wang *et al.* (2019) desenvolveram agrupamento hierárquico com base no mecanismo específico de biotransformação inicial na remoção de princípios ativos por processo de lodo ativado. Os autores verificaram que a eficiência de remoção de princípios ativos pertencentes ao mesmo cluster tende a ser direcionada pelas mesmas variáveis, como coeficiente de partição, massa molecular, tempo de retenção de sólidos e concentração afluente.

Vakarelska *et al.* (2022) aplicaram Análise de cluster hierárquico, Análise de componentes principais, e *Self Organizing Maps* para agrupar micropoluentes em 04 classes com base nas principais características estruturais. Os autores identificaram o potencial de aplicação do método para a extração de padrões para o reconhecimento e investigação de novos contaminantes a partir de propriedades estruturais de moléculas similares.

Baek *et al.* (2021) reduziram os custos de análises para obtenção de padrão *stable isotope labeled (SIL)* propondo um método alternativo para estimar a concentração de micropoluentes em matriz aquática com base em algoritmos de *Deep Learning* e *Machine Learning*. A rede ResNet 101 superou as outras, obtendo R^2 igual a 0,84 e o método de Floresta Aleatória demonstrou R^2 de 0,69, confirmando que a metodologia de análise é uma alternativa assertiva.

Os estudos citados demonstram o potencial na aplicação de métodos de Aprendizado de Máquina para a predição da remoção de poluentes em matrizes aquáticas com base nos parâmetros operacionais das tecnologias aplicadas para remoção. No entanto, a literatura mostra que ao se considerar o potencial de interpretabilidade de modelos, as árvores de decisão têm se destacado entre os principais métodos implementados (HADJIMICHAEL; COMAS; COROMINAS, 2016), sendo base para um conjunto de outras técnicas que apresentam maior potencial preditivo, os métodos *ensemble*.

Ahmed *et al.* (2024) investigaram diferentes métodos baseados em árvores para prever o percentual de adsorção de Ciprofloxacino de água contaminada, a partir da concentração do fármaco (mg/L), dose de adsorvente (mg/L), pH da solução e tempo de adsorção (min). O modelo tradicional de Árvore de Decisão apresentou R^2 de 0,94 e MAE igual a 5,7123, sendo

superado por métodos ensemble como Random Forest ($R^2 = 0,97$, MAE = 3,84) e Gradient Boosting ($R^2 = 0,98$, MAE = 2,62).

Situação semelhante foi identificada por Nguyen et al. (2024), cujo desempenho de Árvores de Decisão ($R^2 < 0,3$) apresentou alto RMSE, para a predição de toxicidade de efluente industrial a partir de temperatura, condutividade, Oxigênio dissolvido, pH, salinidade, concentração de metais tóxicos ($\mu\text{g/L}$). Entre os modelos implementados, as técnicas Adaboost e XGBoost apresentaram maior desempenho ($R^2 = 0,85 - 0,90$).

Outros trabalhos discutem o uso de métodos baseados em árvores no contexto da remoção de poluentes e predição da qualidade de efluentes, demonstrando que esses modelos são capazes de atingir alto desempenho e identificar relações influentes entre as variáveis (CHEN et al., 2025; TAOUIK et al., 2022; ZHU et al., 2021).

Nesse sentido, a próxima subseção introduz conceitos relevantes para a melhor compreensão de métodos baseados em árvores, e estende a apresentação dos conceitos para três técnicas promissoras de *Boosting* utilizadas para predição ou classificação. No caso desta pesquisa, as técnicas serão discutidas à luz de problemas de classificação para métodos de árvores, visto que estes são interpretáveis, lidam bem com valores faltantes, são robustos a *outliers* e são escaláveis (HASTIE; FRIEDMAN; TIBSHIRANI, 2001).

3.3.2 Métodos de classificação baseados em árvores

A classificação é uma abordagem supervisionada para a qual ocorre uma tomada de decisão com base em um padrão prévio. A ideia central é que o método de classificação seja capaz de extrair o padrão dos dados e estabeleça regras para as quais a novos dados x sejam atribuídas K classes pré-definidas ($k = 1, 2, \dots, K$), a partir das suas características de entrada. Assim, estabelece-se uma regra para a qual uma nova observação é identificada e atrelada a uma das classes (rótulos) definidas anteriormente (BISHOP; NASRABADI, 2006; MICHIE *et al.*, 1995).

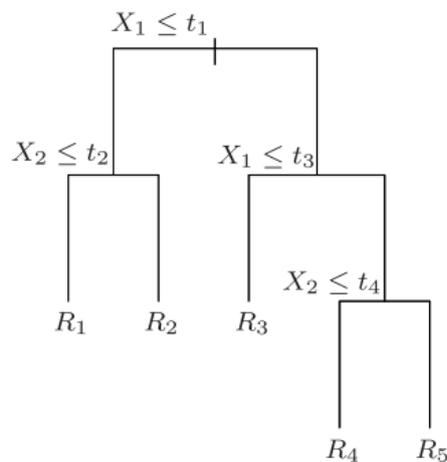
Portanto, a classe é o resultado que será predito a partir dos padrões identificados nas variáveis de entrada (numéricas ou categóricas). A predefinição das classes parte dos atributos de entrada, e com base em um atributo ou mais é definida uma condição de partição e geradas regiões de decisão, que podem ser uma representação binária para duas classes ou multiclasse quando há mais de duas classes (BISHOP; NASRABADI, 2006).

Diferentes métodos de classificação podem ser citados, iniciando por clássicos como Análise Discriminante Linear, Quadrática e Logística, por exemplo (MITCHELL, 1994).

Adicionalmente, uma série de métodos que não assumem uma distribuição prévia para os dados, chamados métodos não-paramétricos, podem ser apresentados como K vizinhos mais próximos (*K-nearest-neighbour*, *K-NN*), *Naive Bayes*, *Multivariate Adaptive Regression Spline (MARS)*, além de outros, incluindo os métodos baseados em árvores (MOLINA; DE LA BLANCA; TAYLOR, 1995).

Nesse sentido, os métodos baseados em árvores são de particular interesse devido ao seu potencial de interpretabilidade. São baseados na separação do espaço de características de entrada em K regiões para as quais um modelo é ajustado em cada uma delas e, as observações atribuídas na mesma região apresentarão a mesma resposta. No caso binário, duas regiões serão divididas recursivamente e um modelo irá prever a melhor resposta Y em cada uma, executando uma abordagem *top-down*. O processo se encerra quando um critério de parada é satisfeito (HASTIE; FRIEDMAN; TIBSHIRANI, 2001). Uma única árvore permite identificar todo o processo de partição realizado, como observado na Árvore de Decisão da Figura 3.

Figura 3 – Partição em Árvore de Decisão



Fonte: Hastie *et al.* 2001.

Identificam-se como variáveis preditoras (características de entrada) X_1 e X_2 , enquanto t_1 , t_2 , t_3 e t_4 são os valores de corte que essas variáveis assumem. Assim, a partição é iniciada pelo topo da árvore pelo nó interno $X_1 \leq t_1$, que gera duas grandes regiões. À esquerda, todas as observações que atendem à condição do nó $X_1 \leq t_1$ e à direita, as observações que atendem a condição em que a variável X_1 apresenta valores maior que t_1 , $X_1 > t_1$. A árvore contém quatro nós internos e cinco nós terminais, ou folhas, indicados pelas regiões R_1 , R_2 , R_3 , R_4 e R_5 . As observações de cada folha apresentam o mesmo resultado. Pode ser deduzido da imagem que a

variável mais importante para a predição é X_1 , visto que ela foi a primeira a ser selecionada pelo algoritmo (JAMES *et al.*, 2023).

Automaticamente, o algoritmo selecionado faz a escolha das variáveis de partição de cada nó a partir de um critério de minimização de erro. No caso da Árvore de Classificação, comumente se considera o Erro de Classificação, Índice Gini ou Desvio, sendo que se espera obter nós terminais com a maior pureza possível. A resposta predita em cada região é a classe majoritária das observações de treino daquela região (HASTIE; FRIEDMAN; TIBSHIRANI, 2001).

É possível que o processo de partição gere predições satisfatórias para os dados de treino, mas não seja eficiente quando aplicado aos dados de teste, levando à construção de árvores profundas e complexas. Esse comportamento é conhecido por *overfitting* ou sobreajuste. Por outro lado, uma árvore pequena com menos regiões de partição pode facilitar a interpretação do processo e ainda, reduzir a variância (JAMES *et al.*, 2023).

Para minimizar o *overfitting*, é possível utilizar estratégias como a poda (*pruning*) ou a regularização, cujo foco é restringir a liberdade da árvore a partir da definição de valores para os parâmetros de ajuste do modelo. Assim, a flexibilidade da árvore é reduzida, e como consequência, a complexidade também (HASTIE; FRIEDMAN; TIBSHIRANI, 2001; JAMES *et al.*, 2023).

O avanço no conhecimento sobre a aplicação de modelos de árvores associado ao potencial computacional possibilitou o desenvolvimento de outras abordagens que visam garantir maior capacidade de generalização aos modelos de árvore construídos.

Nesse sentido, a técnica de *bagging* se caracteriza por tomar diferentes subconjuntos reamostrados dos dados iniciais e, para cada um, construir uma árvore independente e, ao final, agregar os resultados. De forma parecida, o Random Forest melhora o *bagging*, limitando a escolha da variável de corte a um subconjunto dos preditores aleatórios (TAOUFIK *et al.*, 2022).

Uma outra estratégia para melhorar a capacidade preditiva de árvores é o *Boosting*. Nesse caso, as árvores são construídas sequencialmente, a partir da informação residual da árvore anterior. Ou seja, a resposta Y será o próprio resíduo, que será atualizado a cada nova árvore. Com isso, árvores menores são criadas e é possível melhorar o desempenho nas observações classificadas incorretamente. Essa abordagem é poderosa, reduzindo consideravelmente a propensão ao *overfitting* e mantendo a interpretabilidade (JAMES *et al.*, 2023).

Assim, a abordagem *Boosting* visa construir árvores de classificação sequenciais a partir das informações da árvore imediatamente anterior e, ao final, combina o resultado de todas as árvores (JAMES *et al.*, 2023). Esse mecanismo foca no treinamento das observações que anteriormente foram classificadas incorretamente e, possibilita que a cada passo, o erro da classificação seja reduzido. A literatura mostra que modelos *boosting* tem superado outros métodos ensemble no campo de remoção de poluentes em águas (AHMED *et al.*, 2024; NGUYEN *et al.*, 2024).

Yaqub *et al.* (2022) propuseram modelos *ensemble* para prever o percentual de remoção de diclofenaco, iopromida, propranolol e trimethoprima, a partir de pH, Carbono orgânico dissolvido, tempo de operação, dióxido de nitrogênio, sulfatos, nitratos, condutividade elétrica, manganês e ferro. Os modelos XGBoost apresentaram valores de $R^2 > 0,92$ para os fármacos, confirmando seu alto poder de predição e aplicabilidade para mapear os fatores que afetam a remoção dos fármacos e, reduzir a necessidade de análises experimentais.

Chen *et al.* (2025) integraram diferentes grupos de variáveis para desenvolver um modelo capaz e fornecer um índice de qualidade de efluentes. Os autores consideraram parâmetros operacionais, íons metálicos e inorgânicos, parâmetros de biotoxicidade e a concentração de diversos poluentes emergentes, dentre eles, os fármacos tetraciclina, acetaminofeno e gemfibrozil. Dos modelos implementados, a acurácia da Árvore de Classificação foi de 70%, sendo superada pelo XGBoost, LightGBM e CatBoost, que atingiram valores $> 90\%$. O índice de qualidade gerado confirmou que a avaliação assertiva da qualidade de efluentes deve integrar variáveis de diferentes categorias.

Wang *et al.* (2022) propuseram modelos de aprendizado para a predição de Sólidos suspensos totais e Fosfato no efluente de ETE, a partir de parâmetros operacionais como temperatura, vazão, Oxigênio dissolvido, sólidos totais, entre outros. Os autores indicaram *overfitting* do modelo Random Forest, dado que o erro para o conjunto de teste foi significativamente maior do que para o conjunto de treinamento. O modelo XGBoost obteve o melhor desempenho.

3.3.2.1 *Adaptive Boosting* – Adaboost

Proposto por Freund e Schapire (1997), o método Adaboost considera a construção iterativa de árvores de classificação de apenas um nó interno e dois nós terminais, denominadas *stumps*. As predições de cada classificador são combinadas para gerar um resultado final. O método se baseia na atualização de pesos das instâncias de treino a cada etapa realizada, sendo

que na primeira iteração os pesos são idênticos e equivalentes a $1/N$ (tamanho amostral). Na sequência, os pesos são atualizados para o ajuste de uma nova *stump* de modo que as instâncias classificadas incorretamente tenham pesos aumentados por um fator exponencial. Assim, a nova *stump* dará mais influência às observações de treino incorretamente classificadas anteriormente e os pesos são alterados a cada passo.

Esse processo ocorre iterativamente e cada classificador construído terá sua taxa de erro computada e um peso logarítmico que indica a sua contribuição para a predição final. Assim, os classificadores com menor taxa de erro apresentam peso maior. Por fim, os classificadores treinados são combinados por um critério de votos majoritários para gerar a predição final (HASTIE; FRIEDMAN; TIBSHIRANI, 2001). O pseudoalgoritmo do método é exposto no Quadro 3.

Quadro 3 - Pseudoalgoritmo Adaboost

1. Inicialize os pesos das observações $w_i = \frac{1}{N}, i = 1, 2, \dots, N$.

2. Para $m = 1$ até M :

a) Ajuste um classificador $G_m(x)$ aos dados de treinamento usando os pesos w_i .

b) Calcule

$$err_m = \frac{\sum_{i=1}^N w_i I(y_i \neq G_m(x_i))}{\sum_{i=1}^N w_i}.$$

c) Calcule $\alpha_m = \log((1 - err_m)/err_m)$.

d) Defina $w_i \leftarrow w_i \cdot \exp[\alpha_m \cdot I(y_i \neq G_m(x_i))], i = 1, 2, \dots, N$.

3. Resposta $G_x = \text{sign}[\sum_{m=1}^M \alpha_m G_m(x)]$

Fonte: Hastie, Friedman, Tibshirani (2001).

Assim, a resposta final G_x representa a combinação dos classificadores fracos em um único classificador forte. O mecanismo apresentado confirma que a cada iteração, os pesos relativos aos dados classificados incorretamente aumentam, enquanto os pesos das instâncias corretamente classificadas diminuem conforme uma função exponencial.

3.3.2.2 Gradient Boosting Model – GBM

O método Adaboost foi precursor de uma nova classe de modelos rumo à redução da taxa de erro de classificação, apresentando desempenho superior a uma árvore de classificação única (HASTIE; FRIEDMAN; TIBSHIRANI, 2001). Entretanto, é restrito à função de perda Exponencial e por isso, não pode ser considerado flexível.

Nesse sentido, o Modelo de Boosting Gradiente (FRIEDMAN, 2001) se baseia na otimização pelo Gradiente Descendente (*Steepest descent*) para minimização da função de perda. No caso da classificação, aplica-se a função de perda do Desvio binomial (HASTIE; FRIEDMAN; TIBSHIRANI, 2001).

A partir dos resíduos da árvore construída, o GBM identifica os classificadores fracos e adiciona novas árvores de classificação iterativamente, com o objetivo de reduzir o valor da função de perda (TOHARUDIN *et al.*, 2023).

O Gradiente descendente verifica o gradiente local da função de perda dado um conjunto de parâmetros e avança rumo à redução do valor observado. Esse processo ocorre com base na taxa de aprendizado, que define o tamanho do passo de aprendizado do algoritmo.

Os pseudo-resíduos, calculados pela função do gradiente descendente, são utilizados para treinar a *base learner* (Árvore de Classificação) e geram como resultado os nós terminais que indicarão a probabilidade das classes para aquela folha. Com base nessa probabilidade calcula-se novos pseudo-resíduos para cada instância, utilizados para construir a árvore seguinte, e assim sucessivamente até que se atinja o critério de parada. As probabilidades de classe para as folhas de cada árvore serão combinadas para formar a classificação final (HASTIE; FRIEDMAN; TIBSHIRANI, 2001).

O algoritmo paralisa as iterações quando o erro é minimizado de forma consistente ou quando se atinge a quantidade máxima de árvores. Este método é considerado vantajoso visto que é uma abordagem *Boosting* generalizada para diferentes funções de perda e com um vasto conjunto de hiperparâmetros de ajuste-fino, o que o torna mais flexível do que o Adaboost.

No entanto, a possibilidade de ajustes que o algoritmo possui aliado ao fato de que as árvores construídas são mais profundas que no Adaboost, tipicamente torna sua execução mais demorada (TOHARUDIN *et al.*, 2023). No Quadro 4 verifica-se o pseudoalgoritmo desse método.

Quadro 4 - Pseudoalgoritmo Gradient Boosting Model

1. Inicialize $f_0(x) = \arg \min_{\gamma} \sum_{i=1}^N L(y_i, \gamma)$.

2. Para $m = 1$ até M :

a) Para $i = 1, 2, \dots, N$ calcule

$$r_{im} = - \left[\frac{\partial L(y_i f(x_i))}{\partial f(x_i)} \right]_{f = f_{m-1}}$$

b) Ajuste uma árvore de regressão aos alvos r_{im} com regiões terminais $R_{jm}, j = 1, 2, \dots, J_m$.

c) Para $j = 1, 2, \dots, J_m$ calcule

$$\gamma_{jm} = \arg \min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma).$$

$$d) \text{ Atualize } f_m(x) = f_{m-1}(x) + \sum_{j=1}^m \gamma_{jm} I(x \in R_{jm}).$$

$$3. \text{ Resposta } \hat{f}(x) = f_M(x)$$

Fonte: Hastie, Friedman, Tibshirani (2001).

3.3.2.3 Extreme Gradient Boosting – XGBoost

O método XGBoost foi proposto por Chen e Guestrin (2016) como uma otimização ao GBM. O algoritmo apresenta um termo de regularização avançada que previne o *overfitting* e otimiza a função de perda pelo método de Newton, o que garante rápida convergência. Além disso, pode realizar validação cruzada a cada passo iterativo e ainda efetuar a parada da execução ao atingir uma quantidade ótima de iterações.

É reconhecido por apresentar alta velocidade em sua execução devido ao fato de apresentar processamento paralelo. O algoritmo também é altamente flexível devido a vasta quantidade de hiperparâmetros para ajuste. Nesse sentido, a função objetivo desse algoritmo apresenta o termo de perda adicionado a um termo de regularização. Pode ser representada por:

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad \text{Eq. (1)}$$

Sendo que y_i é o i -ésimo rótulo observado, \hat{y}_i é o i -ésimo rótulo predito pela árvore anterior, l é a função de perda, f_t é a saída da t -ésima árvore, a regularização $\Omega(f_t)$ é indicada pela Eq. (2).

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad \text{Eq. (2)}$$

Sendo T o número de folhas, ω os scores das folhas, γ e λ são hiperparâmetros de perda e regularização. O *score* ou peso da folha é calculado de acordo com a Eq. (3).

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad \text{Eq. (3)}$$

Os termos (g_i) e (h_i) resultam da solução do termo de perda da função objetivo pela aproximação de Taylor, e representam gradientes da perda residual. Considerando manipulações na Eq. (1) e a expansão do termo de regularização, a função objetivo simplificada passa a ser representada pela Eq. (4).

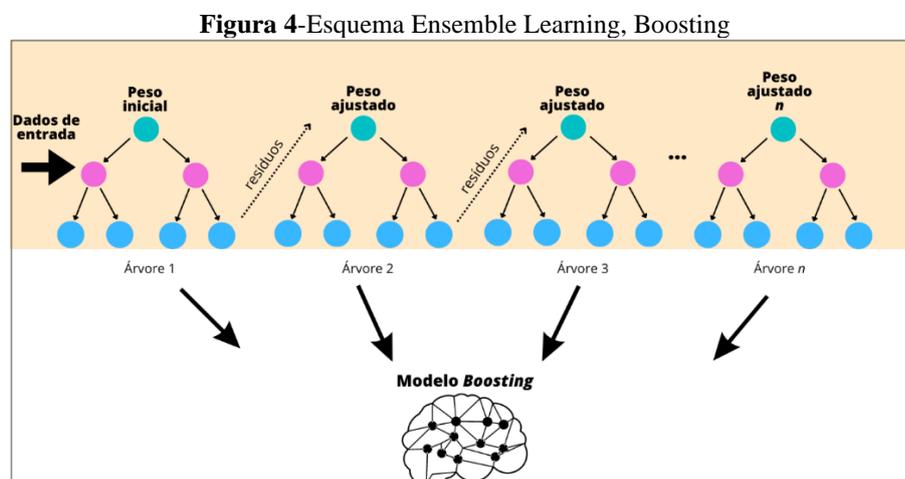
$$Obj^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad \text{Eq. (4)}$$

O processo de crescimento da árvore se inicia com uma única folha, como predição inicial. A partir dos resíduos obtidos, cria-se a primeira árvore do *ensemble*, onde para cada nó será testada uma nova separação. Assim, uma nova partição só será adicionada se, de fato, proporcionar ganho. Essa avaliação é realizada de acordo com a Eq. (5).

$$ganho = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad \text{Eq. (5)}$$

Sendo que G e H representam o somatório dos gradientes residuais g_i e h_i para as folhas à esquerda (L) e à direita (R).

Por fim, uma sequência de modelos preditivos é chamada de *Ensemble* e quando o aprendizado ocorre a partir do agrupamento de modelos diz-se Aprendizado por Agrupamento, ou *Ensemble Learning*. Os métodos que executam o aprendizado por agrupamento de modelos preditivos são chamados de Métodos *Ensemble*, como é o caso dos métodos *Boosting*, cujo esquema geral é apresentado na Figura 4.



Fonte: Elaborado pela autora.

Na Figura 4 verifica-se que os dados de entrada de cada árvore provêm do resultado da árvore anterior, de forma que as observações apresentam pesos diferentes a cada árvore construída. Ao final, os resultados das árvores construídas serão combinados para originar um resultado único de classificação, com base na classe majoritária.

3.3.3 Métricas de avaliação de modelos de classificação

Após implementar um algoritmo é necessário avaliar o desempenho apresentado pelo modelo. Para isso, diferentes métricas de avaliação de desempenho podem ser utilizadas. No caso de classificação, tipicamente, considera-se a Matriz de Confusão, Precisão, *Recall* e *F1 - score* (GÉRON, 2022).

A Matriz de Confusão é uma forma de avaliação em que se apresenta a frequência em que uma observação de uma classe é classificada erroneamente em outra classe. Esse formato de avaliação é gerado a partir das predições, que são comparadas com os valores reais. Cada linha da matriz corresponde a uma classe real e cada coluna corresponde a predição das classes. É possível apresentar a Matriz de Confusão como uma matriz de valores ou, de forma mais visual através de um mapa de calor. Na Figura 5 verifica-se uma matriz de confusão binária. Na figura, as classes são representadas como 0 (classe negativa) e 1 (classe positiva).

Figura 5 – Matriz de Confusão

Valor real	Negativo (0)	VN	FP
	Positivo (1)	FN	VP
		Negativo (0)	Positivo (1)
		Valor predito	

Fonte: Elaborado pela autora.

A sigla VP corresponde aos verdadeiros positivos, FP aos falsos positivos, FN aos falsos negativos e VN aos verdadeiros negativos.

Outra métrica amplamente utilizada é a precisão, que corresponde a fração de predições positivas que realmente pertencem à classe positiva. Essa métrica pode ser obtida a partir da Matriz de Confusão, de acordo com a Equação (6).

$$\frac{VP}{VP + FP} \quad \text{Eq. (6)}$$

Sendo assim, a métrica traz a proporção de acertos considerando as predições totais da classe considerada positiva. É utilizada em conjunto com o *Recall* ou Sensibilidade, que indica a taxa de acertos entre os positivos reais, ou seja, as observações positivas corretamente identificadas pelo classificador a partir da Equação (7).

$$\frac{VP}{VP + FN} \quad \text{Eq. (7)}$$

O *F1 - score* é uma combinação entre precisão e *Recall*, a partir da média harmônica dos valores de forma que o valor de *F1 - score* só será alto se realmente ambos os valores forem altos. É uma forma simples de se comparar dois classificadores, como indicado na Equação (8) (GÉRON, 2022).

$$F1 = 2 \times \frac{\text{precisão} \times \text{recall}}{\text{precisão} + \text{recall}} \quad \text{Eq. (8)}$$

A acurácia também é uma métrica de interesse, utilizada para avaliar o desempenho geral de modelos de classificação. Indica os acertos do modelo de forma global (tanto positivos quanto negativos), como indicado na Equação (9).

$$\frac{VP + VN}{VP + FP + FN + VN} \quad \text{Eq. (9)}$$

Os valores de Precisão, *Recall* e *F1 - score* variam entre [0 e 1] e quanto maior o valor, melhor é o desempenho do modelo.

3.3.4 Explicabilidade de modelos de aprendizagem: método SHAP

A interpretabilidade pode ser entendida como o nível ao qual um humano consegue prever o resultado do modelo. Nesse caso, quanto mais interpretável for um modelo de aprendizado, mais fácil se torna a compreensão dos motivos que levaram às decisões do modelo,

ou predições. Exemplos de modelos interpretáveis são Regressão Linear e Árvores de Decisão. No entanto, o aumento da complexidade dos modelos e da quantidade de parâmetros impede a compreensão profunda dos resultados. Esse é o caso dos métodos Boosting.

Para melhor compreender os resultados do modelo, torna-se necessário utilizar técnicas de explicabilidade capazes de trazer *insights* sobre as decisões do modelo, considerando tanto as relações existentes nos dados como as aprendidas pelo modelo. Assim, entender o motivo pelo qual uma decisão foi tomada auxilia a uma melhor compreensão do problema investigado, dos dados utilizados ou, ainda, razões pelas quais o modelo poderia vir a falhar. Essa compreensão aumenta a confiança nos resultados obtidos e, facilita a aceitabilidade social do uso de modelos de aprendizado (MOLNAR, 2022).

Uma vez que diferentes modelos podem ser testados para um mesmo problema, é de interesse escolher livremente os modelos, mas empregar métodos complementares para adicionar explicabilidade aos resultados.

O SHAP (*SHapley Additive exPlanations*) (LUNDBERG, 2017) é um método para a explicação de predições, a partir do cálculo de valores Shapley advindos da teoria dos jogos. Esse cálculo possibilita verificar o quanto cada variável contribui para a predição final da instância considerada. Ou seja, os *Shapley values* representam o efeito da inclusão das variáveis na predição individual do modelo para cada instância, com base na Eq. (10).

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad \text{Eq. (10)}$$

Sendo F o conjunto de todas as variáveis do modelo, treina-se tanto um modelo com a variável presente, como um modelo sem a variável. Assim, S é o subconjunto de variáveis consideradas para investigar o efeito da inclusão de cada uma delas no modelo.

O método SHAP, pode ser considerado uma complementação dessa abordagem, por permitir a verificação do impacto de cada variável no modelo global pela agregação dos *Shapley values*, como indicado na Eq. (11).

$$g(x') = \phi_0 + \sum_{i=1}^M \phi_i x'_i \quad \text{Eq. (11)}$$

Sendo $g(x')$ o modelo de explicação, x' o vetor de combinação das variáveis ou variável individual a ter sua influência investigada, M a quantidade de combinações, e ϕ_i os

Shapley values para uma variável i . Desse modo, o SHAP garante interpretação local e global, além de expressar a importância das variáveis, dependência das variáveis e, interações (LUNDBERG, 2017).

As relações identificadas podem ser representadas de forma visual e intuitiva, o que facilita a interpretação dos resultados alcançados pelos modelos de aprendizado e, proporcionam ganho de conhecimento acerca dos problemas reais investigados. Devido a isso, justifica-se a aplicação do método SHAP para a interpretação de modelos de classificação *Boosting* (OLIVEIRA; SANT'ANNA, 2023; TOHARUDIN *et al.*, 2023).

3.5 INOVAÇÃO DA PROPOSTA

A revisão de literatura confirma o substancial desafio para gerenciar a ocorrência de fármacos nas matrizes aquáticas, dadas as limitações financeiras e tecnológicas para adquirir dados da ocorrência ambiental dos compostos. Frente a isso, métodos de aprendizado de máquina podem ser uma alternativa viável que para reduzir a necessidade frequente de coleta de novos dados, ao passo que os modelos gerados podem apresentar alto potencial de adaptação e implementação.

A abordagem proposta é inovadora, visto que utiliza modelos de aprendizado a partir de dados de comercialização dos fármacos e seu percentual de remoção na ETE. A associação concomitante de dados dessas duas naturezas para a predição final do nível de ocorrência dos compostos ainda não foi avaliada na literatura científica. Ademais, os dados de comercialização ainda pouco explorados na literatura também reforçam a inovação desta pesquisa.

Por fim, o mecanismo proposto pode ser uma referência para novas adaptações futuras.

4 METODOLOGIA

4.1 DESCRIÇÃO GERAL DA PESQUISA

Os modelos preditivos tipicamente aplicados para estimar a concentração de fármacos no efluente tratado de esgoto a partir de registros de venda de fármacos se baseiam em uma metodologia consolidada pela *European Medicine Agency*, vastamente aplicada a partir da inclusão do Fator de Excreção Humano (HE *et al.*, 2020; VERLICCHI *et al.*, 2014). Essa metodologia considera que a concentração afluenta de fármaco é dependente da quantidade anual de venda de fármacos, população da área investigada, volume per capita de esgoto e, quando adaptada, o fator de excreção humano. Para concentração efluente, multiplica-se o resultado anterior pela eficiência de remoção dos fármacos em ETE. No entanto, essa abordagem tradicional não garante o aprendizado, ou seja, a capacidade de que o modelo se adapte a diferentes padrões de entrada e, não considera outras variáveis relevantes para os níveis ambientais dos fármacos.

Portanto, o uso de métodos de Aprendizado de Máquina para estimativa da ocorrência de poluentes ou da remoção desses compostos em matrizes aquáticas tem mostrado alto potencial nos estudos de meio ambiente, como discutido por Taoufik *et al.* (2022) e Hadjimichael *et al.* (2016), e resultam em bom desempenho e, devido a capacidade de aprendizado, se adaptam aos novos dados. Os modelos de aprendizado têm mostrado que parâmetros operacionais fornecem insights relevantes para a compreensão dos mecanismos de remoção dos fármacos. No entanto, até a presente data, esses modelos não integraram os dados de comercialização dos fármacos. Dessa forma, cresce o interesse em investigar o potencial de modelos de Aprendizado de Máquina para prever a concentração final de fármacos no efluente de ETE, mas considerando as vendas de fármacos como relevante variável de entrada.

Nesta pesquisa, abordou-se o problema a partir da investigação de métodos *Boosting*, que se baseiam no modelo de Árvore de Classificação, ambos descritos no item 3.3.2. As variáveis consideradas para a construção dos modelos foram a quantidade de fármacos comercializada, o fator de excreção humano de cada fármaco, eficiência de remoção dos fármacos em ETE, a concentração afluenta de fármacos em ETE e a concentração efluente de fármacos em ETE.

Para tanto, propuseram-se modelos baseados em dados para classificar a concentração de ocorrência de fármacos ao final do processo de tratamento em ETE, levando em

consideração a quantidade de fármacos vendida à população. Portanto, a variável resposta do modelo é a classe da concentração final de fármaco obtida a partir de valores de concentração de saída do processo de tratamento em ETE. Com base na literatura (HE et al., 2020; VERLICCHI et al., 2014) e nos dados disponíveis, considerou-se como variáveis explicativas a concentração de entrada de princípio ativo com base na quantidade de fármacos comercializada (PEC_{inf}), unidades de fármacos vendidos (caixas ou frascos), quantidade contida em uma unidade vendida (número de comprimidos ou volume de frasco), massa total comercializada por ano ($kg \cdot ano^{-1}$), Fator de Excreção Humano e Eficiência de remoção no tratamento em ETE.

Os dados da comercialização de fármacos foram obtidos do Sistema Nacional de Gerenciamento de Produtos Controlados (SNGPC) disponibilizado pela Agência Nacional de Vigilância Sanitária (ANVISA) no Portal Brasileiro de Dados Abertos, os dados relativos ao processo de tratamento em ETE foram coletados em experimentos reportados na literatura e a partir do Atlas Esgotos publicado pela Agência Nacional de Águas e Saneamento Básico (ANA, 2017).

A base de dados abertos contendo registros da comercialização de medicamentos industrializados sujeitos à escrituração no SNGPC tem por finalidade controlar os trâmites de compra, venda, transferências ou perdas de medicamentos comercializados em municípios brasileiros. Por isso, farmácias e drogarias privadas atribuíram periodicamente os dados de comercialização para processamento e armazenamento no sistema, até dezembro de 2021.

Assim, medicamentos controlados, antimicrobianos, medicamentos que contenham cloroquina, hidroxicloroquina, nitazoxanida ou ivermectina foram de escrituração obrigatória, de acordo com a Portaria SVS/MS nº 344/1998 (ANVISA, 1998), Resolução RDC nº 20/2011 (ANVISA, 2011) e atualizações (ANVISA, 2017; 2021; 2014). Entretanto, não houve impedimento na declaração de vendas de outros medicamentos (PORTAL DE DADOS ABERTOS, [s.d.]).

No SNGPC, reportam-se dados da dispensação de fármacos para 16 variáveis: ano, mês, venda, estado, município, princípio ativo, descrição de apresentação, quantidade vendida, unidade de medida, conselho prescriptor, estado do conselho prescriptor, tipo de receituário, CID-10, sexo, idade e unidade da idade. Cada observação reportada na base corresponde a uma ocorrência de venda. A descrição das variáveis pode ser verificada no ANEXO A – Dicionário de dados abertos da ANVISA.

Paralelamente, o “Atlas Esgotos: Despoluição de Bacias Hidrográficas” é parte de uma estratégia desenvolvida pela ANA e Secretaria Nacional de Saneamento Ambiental do

Ministério das Cidades (SNSA/MCidades) que visa melhor compreender o cenário de esgotamento sanitário dos municípios brasileiros, considerando o impacto do lançamento de cargas efluentes nos corpos hídricos receptores. Além disso, constitui um documento que auxilia a proposição de estratégias de gerenciamento de esgotos em relação ao tratamento, proteção de recursos hídricos e aporte de recursos financeiros (ANA, 2017).

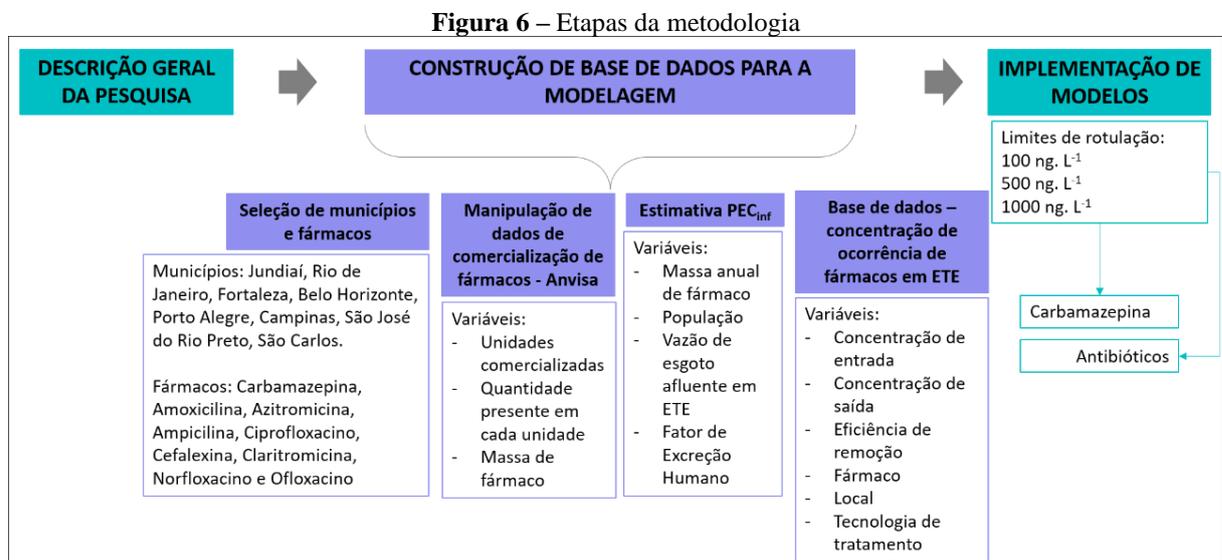
O Atlas Esgotos apresenta resultados de investigações realizadas em cada município brasileiro, em parceria com as prestadoras de serviço de saneamento. Dessa forma, os 5.570 municípios foram segregados em dois grupos distintos para o levantamento de dados:

Grupo 1: Municípios que possuem prestadores de serviço de saneamento institucionalizados e municípios que não apresentam prestadores de serviço institucionalizados, mas possuem população maior do que 50.000 habitantes. Este grupo soma 3005 municípios.

Grupo 2: Compreende os municípios com população menor do que 50.000 habitantes e não possuem prestadores de serviço de saneamento institucionalizados. Este grupo soma 2565 municípios.

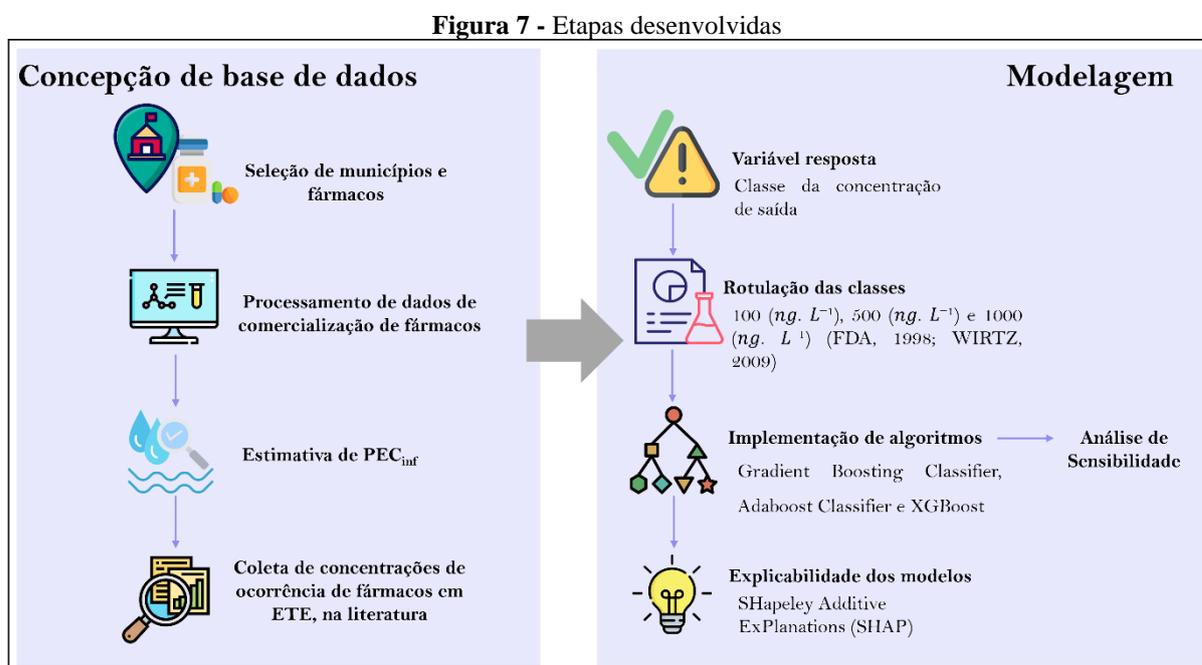
Nesse sentido, para os municípios do Grupo 1 foram coletados dados primários a partir da interação com os prestadores e companhias de serviço. Para o Grupo 2, foram obtidos dados secundários a partir das fontes: Pesquisa Nacional do Saneamento Básico (PNSB), Pesquisa Nacional por Amostra de Domicílios (PNAD) e Sistema Nacional de Informações sobre Saneamento (SNIS) (IBGE, 2002, 2015; MINISTÉRIO DAS CIDADES, 2017).

Na Figura 6 são apresentadas as etapas descritas no presente capítulo.



Fonte: Elaborado pela autora.

Dessa forma, a segunda sessão dos métodos apresenta a concepção da base de dados da qual as variáveis explicativas e variável resposta foram extraídas para o treinamento dos modelos propostos. A base foi concebida pela combinação de dados provenientes da coleta, tratamento de dados e estimativas com base nos registros de venda reportados pela ANVISA associados aos dados de concentrações de ocorrência de fármacos coletados na literatura. Na terceira sessão apresenta-se os detalhes para a construção dos modelos. De modo complementar, na Figura 7 é possível verificar resumidamente todas as etapas realizadas nesta pesquisa.



Fonte: Elaborado pela autora.

Sendo assim, a fase inicial contou com a concepção da base de dados, e posteriormente, a etapa de modelagem foi desenvolvida. Cada etapa representada na Figura 7 será descrita em detalhes nas próximas seções.

4.2 CONSTRUÇÃO DA BASE DE DADOS PARA O MODELO

4.2.1. Seleção de municípios representativos e fármacos comercializados - ANVISA

A seleção dos municípios de interesse partiu dos critérios de que (I) os municípios foram alvo de estudos de monitoramento de fármacos em matrizes aquáticas e (II) fazem parte do

Grupo 1 do Atlas Esgotos. Sendo assim, os municípios de Jundiaí, São José do Rio Preto, Rio de Janeiro, Belo Horizonte, Fortaleza, São Carlos, Porto Alegre e Campinas foram escolhidos como representativos para esta investigação.

Os CPEs da classe dos fármacos foram selecionados para este estudo com base na taxa de comercialização para a população brasileira, antecedentes de identificação e quantificação em ETE e águas superficiais verificados na literatura científica, bem como histórico de inclusão em listas prioritárias de monitoramento. Os fármacos representativos considerados foram os antibióticos Ampicilina (AMP), Amoxicilina (AMOX), Azitromicina (AZT), Cefalexina (CEF), Ciprofloxacino (CIP), Claritromicina (CLA), Norfloxacino (NOR), Ofloxacino (OFL), juntamente com o psicoativo Carbamazepina (CMZ).

Portanto, esta pesquisa considerou a ocorrência dos fármacos selecionados em esgoto sanitário, visto que as ETE são a principal rota de inserção desses compostos em águas superficiais. Os fármacos selecionados são amplamente reportados em publicações científicas, compõem a Lista de prioridade 2020/1161 (EUROPEAN COMMISSION, 2020) e são indicados na literatura científica como prioritários ao monitoramento em matriz aquática (DE VOOGT *et al.*, 2009; KOSTICH; BATT; LAZORCHAK, 2014; REICHERT *et al.*, 2019), como indicado na Tabela 2.

Tabela 2 – Classe prioritária dos Princípios Ativos considerados

Classe Prioritária	Fármaco
Classe I: <i>High priority pharmaceuticals</i>	Carbamazepina Ciprofloxacino
Classe II: <i>Priority Pharmaceuticals</i>	Amoxicilina Ofloxacino Claritromicina
Classe III: <i>Lower priority pharmaceuticals</i>	Cefalexina

Fonte: De Voogt (2009).

Os fármacos Norfloxacino e Ampicilina também são indicados por Reichert *et al.* (2019) como prioritários, bem como a Azitromicina foi incluída na lista de prioridade da União Europeia 2015/495 (EUROPEAN COMMISSION, 2015). Portanto, a preocupação em relação à ocorrência desses contaminantes em matrizes ambientais destaca a relevância e a urgência para investigações que auxiliem no seu monitoramento.

4.2.2. Pré-processamento de dados – ANVISA

Os dados coletados para esta pesquisa contemplam a dispensação mensal de fármacos para o período de 2014 – 2020 para os municípios representativos selecionados. A base de dados da ANVISA contém ao menos 500 insumos farmacêuticos ativos combinados ou em forma de monodrogas (apenas um princípio ativo).

Os dados foram coletados no Portal Brasileiro de Dados Abertos (ANVISA, 2020) por meio do download de planilhas em formato .csv para cada mês, para o período de janeiro de 2014 a dezembro de 2020. Dessa forma, os dados passaram por uma etapa de pré-processamento no software *Rstudio*, utilizando os pacotes disponíveis na coleção *Tidyverse*.

Após a coleta, as observações das planilhas foram agregadas em um mesmo arquivo de dados no *Rstudio* e, ao mesmo tempo, foi aplicado um filtro para selecionar apenas os municípios representativos. A coluna “CID10” foi excluída nesta etapa.

Na base de dados, foi realizada uma uniformização dos nomes dos princípios ativos (fármacos) registrados na coluna “PRINCIPIO_ATIVO”. Por exemplo, os termos “tri hidratado” e “tri-hidratado” foram alterados para “trihidratado”. O mesmo ocorreu com os termos relativos a “dihidratado”. Também foi realizado um processo de exclusão de observações faltantes (NA) da variável “PRINCIPIO_ATIVO”.

Nesse novo arquivo, a taxa de NA foi verificada para cada coluna presente e, também, foram observados valores aberrantes na coluna “IDADE”. Nesta etapa, o arquivo de dados único foi segregado em 6 arquivos distintos correspondentes a cada ano. Portanto, 6 arquivos no formato *RData* foram gerados a fim de viabilizar a velocidade da manipulação dos dados.

Dessa forma, o pré-tratamento foi realizado de maneira semelhante para todos os arquivos. Inicialmente, foi realizado um processo de seleção de observações para os fármacos de interesse, ambos monodrogas: Ampicilina, Amoxicilina, Azitromicina, Cefalexina, Ciprofloxacino, Claritromicina, Norfloxacino, Ofloxacino e Carbamazepina. As formas “monoidratada”, “dihidratada”, “trihidratada”, com nomenclatura previamente padronizada, foram consideradas para Amoxicilina, Azitromicina, Cefalexina e Ampicilina.

Posteriormente, a manipulação da coluna “DESCRICA0_APRESENTA0AO” foi realizada a fim de segregar a descrição do fármaco, visto que esta variável traz dados para o cálculo da massa anual comercializada. A apresentação do fármaco é descrita com base em termos padronizados de acordo com o indicado pelo Vocabulário Controlado da ANVISA (ANVISA, 2011). Verifica-se no Quadro 5 dois exemplos de descrição:

Quadro 5 – Apresentação de medicamentos

DESCRICA0_APRESENTA0AO	Interpretação
------------------------	---------------

5 MG COM CT BL AL X 20	Caixa com 1 blister de alumínio com 20 comprimidos, cada comprimido com 5 mg de princípio ativo
15MG/ML XPE CT FR VD x 50 ML	Caixa com 1 frasco de vidro com 50 mL de um xarope, com concentração do princípio ativo de 15 mg por mL

Fonte: ANVISA (2020, p. 7).

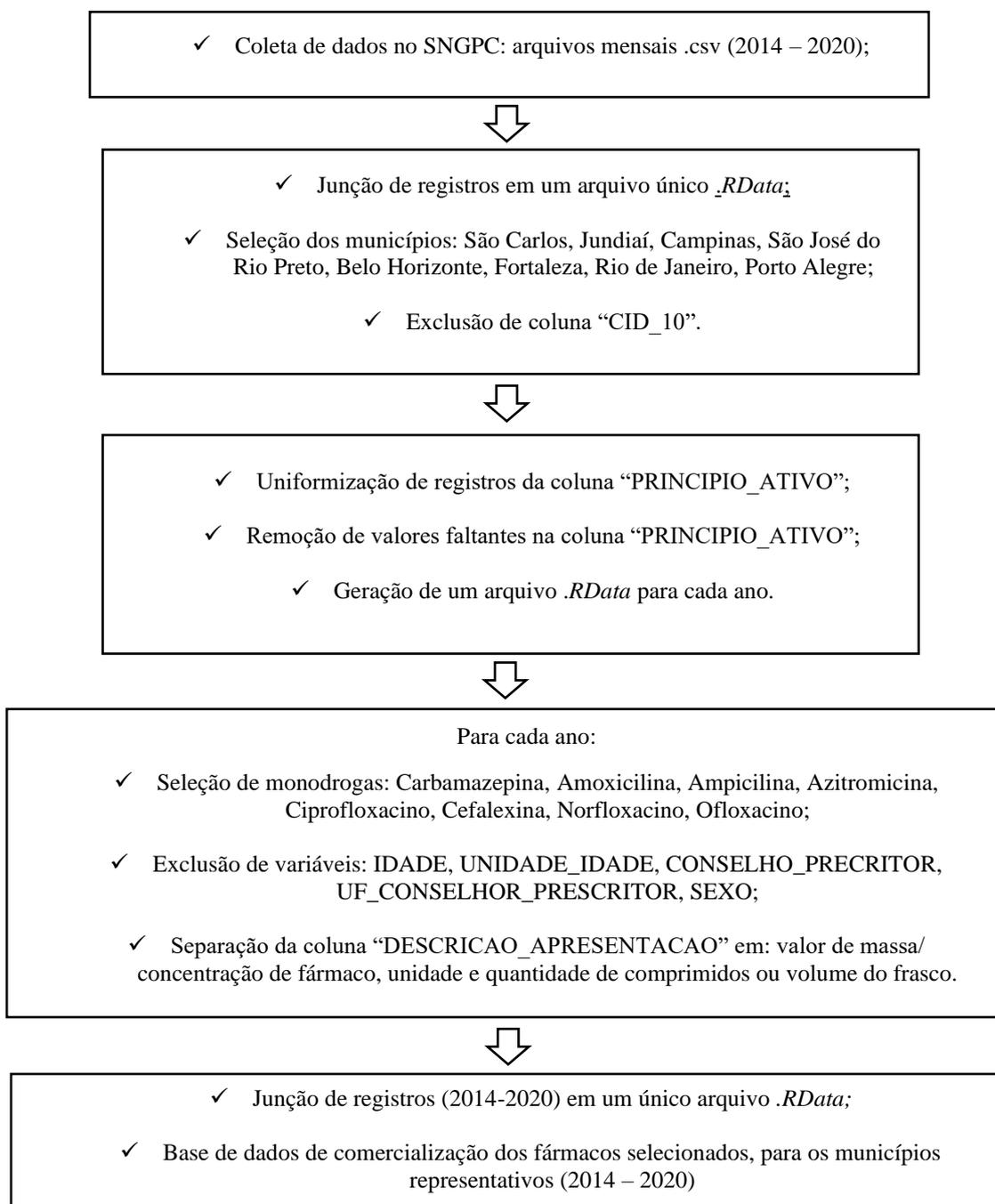
Assim, a coluna de descrição foi padronizada devido a incompatibilidade na apresentação de unidades e, em seguida, foi segregada de forma a obter separadamente a massa ou concentração de princípio ativo, a unidade de medida indicada e a quantidade de comprimidos ou volume total do frasco, no caso de forma líquida. Para tanto, foram identificadas incompatibilidades na forma de registro das unidades de concentração. Por exemplo, unidades como “MG/ ML” ou “MG / ML” foram padronizadas para “MG/ML”. Da mesma forma, “MG/ 5ML”, “MG / 5 ML” e “MG/5 ML” foram padronizadas para “MG/5ML”.

Adicionalmente, foram verificadas observações para as quais a unidade estava conectada sem espaço ao valor de massa ou concentração. Portanto, foi realizada a adição de espaço para posteriormente ser possível separar os registros corretamente. E, por fim, valores de descrição apresentados em gramas foram convertidos para miligramas (mg) de forma a padronizar as etapas de cálculo aplicadas posteriormente.

Desse modo, a coluna de descrição foi excluída da base e as três novas variáveis foram incluídas. Também foram excluídas as colunas idade, tipo de receituário, conselho prescritor, unidade idade e sexo, já que não seriam utilizadas nas etapas posteriores.

O processo descrito foi realizado para os arquivos de dados de cada ano e, ao final, os dados já manipulados e filtrados foram agregados em um mesmo arquivo *.RData*. Portanto, a base de dados final obtida considerou registros de comercialização de monodrogas da classe terapêutica dos antibióticos (Amoxicilina, Claritromicina, Azitromicina, Cefalexina, Ampicilina, Ofloxacino, Norfloxacino, Ciprofloxacino) e droga psiquiátrica (Carbamazepina) para os municípios representativos. A sequência de etapas de pré-processamento dos dados de comercialização de fármacos é indicada no diagrama apresentado na Figura 8.

Figura 8 – Pré-processamento de dados de comercialização de fármacos



Fonte: Elaborado pela autora.

4.2.3. Estimativa de PEC_{inf} com base na comercialização de fármacos

Os dados de comercialização são uma importante forma alternativa para estimar a concentração de ocorrência de fármacos excretados nos esgotos sanitários que são encaminhados para as ETE. Essa concentração é identificada como *Predicted Environmental*

Concentration in wastewater influent ($PEC_{inf, EMA}$). Segundo as diretrizes da *European Medicine Agency* (2006), a $PEC_{inf, EMA}$ é calculada de acordo com a Equação (12):

$$PEC_{inf, EMA} = \frac{A}{365 \times P \times V} \quad \text{Eq. (12)}$$

no entanto, acrescentou-se o Fator de Excreção Humano (f) para estimar PEC_{inf} . Na Equação (13) verifica-se a adaptação proposta por He *et al.*, (2020).

$$PEC_{inf} = \frac{A}{365 \times P \times V} \times f \quad \text{Eq. (13)}$$

Nesse caso, a variável “A” corresponde a quantidade de unidades de fármacos vendida anualmente (kg. ano⁻¹/ ano), a variável “V” equivale ao volume de esgotos gerado por pessoa e por dia (m³) obtido pelos dados da Agência Nacional de Águas (ANA, 2017), a variável “P” indica a população local (ANA, 2017) e o Fator de excreção (f) indica a parcela de fármaco excretada, não metabolizada pelo organismo humano. Esse último foi obtido para cada fármaco a partir da literatura científica (HE *et al.*, 2020; VERLICCHI *et al.*, 2014). Contudo, é possível que haja variações entre os valores reportados pelos autores. Portanto, respeitando o critério proposto por He *et al.* (2020) utilizou-se o maior valor reportado a fim de contabilizar o caso mais extremo.

É importante evidenciar que essa estimativa, embora consolidada, apresenta incertezas. Os dados de comercialização estão suscetíveis à inacurácias, visto que é comum que autores utilizem dados de nível nacional, sendo que o padrão de consumo varia entre regiões de um mesmo país. Por isso, esta pesquisa utiliza dados de comercialização especificamente para os municípios de interesse.

Contabiliza-se, nesta pesquisa, a comercialização dos fármacos de interesse, a partir de dados provenientes apenas de farmácias e drogarias privadas. Portanto, não se considera a potencial dispensação por outros meios. Além disso, pressupõe-se que toda a quantidade comercializada é consumida.

Assim, o percentual de fármacos disposto inadequadamente no lixo comum residencial, no vaso sanitário, ou a parcela retornada às drogarias não são contabilizados. Outra variável relevante é o Fator de Excreção Humano, pois depende diretamente das características do metabolismo dos indivíduos como idade, gênero, consumo de fármacos, entre outras

características. A depender do fármaco, podem ser apontados diferentes valores na literatura e, sugere-se a utilização do valor mais alto como estratégia.

Observa-se na Tabela 3, os valores considerados para o Fator de Excreção Humano.

Tabela 3 – Fator de Excreção Humano

Fármaco	Valor (%)	Referência
Carbamazepina	30	Ternes and Joss (2006) ¹
Ciprofloxacino	64	Zhao <i>et al.</i> (2014)
Amoxicilina	60	RxList ¹
Ofloxacino	90	Al-Omar (2009)
Clarithromicina	20	Kümmerer e Henninger (2003) ²
Cefalexina	20	Dave e Moris (2016)
Norfloxacino	70	Monteiro e Boxall (2010) ¹
Ampicilina	40	Wishart <i>et al.</i> (2017) ²
Azitromicina	14	RxList ¹

¹Valores utilizados em Verlicchi *et al.* (2014). ²Valores utilizados por He *et al.* (2020).

A título de exemplo, o cálculo da massa anual (kg) de fármacos comercializada é detalhado na Tabela 4 e se baseia na descrição do fármaco prescrito, a exemplo do Quadro 5. Nesse caso, considerou-se as vendas de Carbamazepina no ano de 2014, no município Rio de Janeiro. Verifica-se que foram considerados medicamentos sólidos, em cápsulas ou comprimidos, e medicamentos líquidos, com diferentes descrições de apresentação.

Tabela 4 – Cálculo de Massa total comercializada (kg) de Carbamazepina, Rio de Janeiro, 2014.

Massa de Princípio Ativo (mg)	Quantidade total de comprimidos em cada embalagem	Unidades totais comercializadas	Massa comercializada (kg)
20 *	100*	11.980	23,960
200	20	62.596	250,384
200	30	18.809	112,854
200	60	23.692	284,304
400	10	33	0,132
400	20	18.781	150,248
400	60	5.513	132,312
400	200	1	0,08

*A prescrição corresponde ao medicamento de 20 mg/ml, com volume total de 100 ml no frasco.

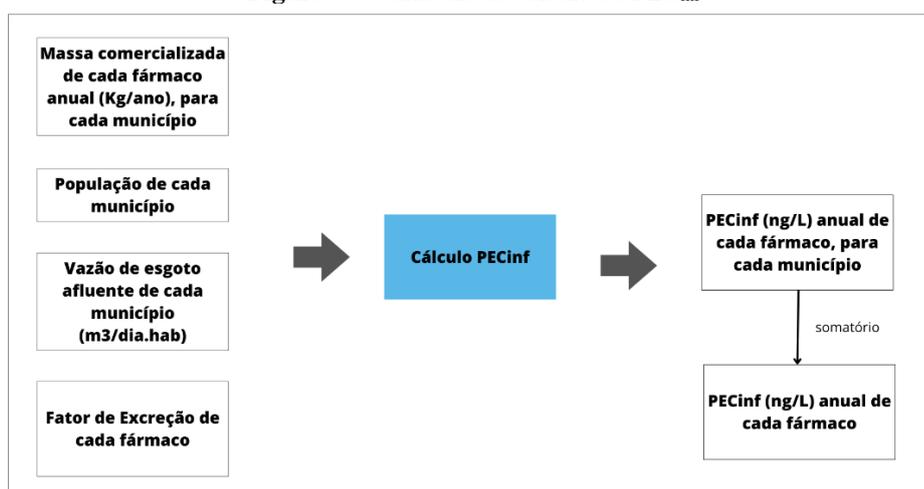
Fonte: Elaborado pela autora a partir de dados da ANVISA.

Assim, a massa total de Carbamazepina, comercializada no ano de 2014, para a cidade do Rio de Janeiro equivale a 954,27 kg. Considera-se que o fator de excreção humano desse fármaco é 30%, o valor de $V = 0,127 \text{ m}^3 \cdot \text{dia}^{-1} \cdot \text{hab}^{-1}$ e $P = 6.429.923$ habitantes, de acordo com a Equação 13 o valor de PEC_{inf} é 959,48 ng. L⁻¹.

A concentração anual total de Carbamazepina para o ano de 2014 avaliada nesta pesquisa, foi obtida através do somatório das parcelas de PEC_{inf} para cada um dos municípios considerados utilizando a mesma rotina de cálculo exemplificada na Tabela 4.

Portanto, de acordo com a quantidade de Carbamazepina comercializada em 2014 nos oito municípios de interesse, uma concentração total afluyente de 8.428,661 ng. L⁻¹ deste princípio ativo chega à ETE para tratamento. Dessa forma, esse cálculo foi realizado para cada ano considerado e para cada fármaco de interesse. A Figura 9 ilustra o processo realizado.

Figura 9 – Variáveis do cálculo de PEC_{inf}



Fonte: Elaborado pela autora.

Considerou-se que a quantidade de fármaco comercializada foi completamente consumida pelo paciente, ou seja, se uma caixa contém 20 comprimidos, admitiu-se que todos foram consumidos. Os valores resultantes do cálculo, foram agregados na base de dados de fármacos apresentada na sessão anterior.

O procedimento descrito nessa sessão foi reproduzido para os municípios de São Carlos, Jundiaí, Campinas, São José do Rio Preto, Belo Horizonte, Fortaleza, Rio de Janeiro e Porto Alegre.

4.2.4. Coleta de concentrações de ocorrência de fármacos nos esgotos sanitários

Paralelamente às manipulações dos dados de comercialização de fármacos, realizou-se uma pesquisa bibliográfica por artigos, teses ou dissertações para obter dados experimentais do processo de remoção de fármacos em esgotos sanitários. As ETE são a principal fonte de emissão de concentrações traço dos princípios ativos farmacêuticos nas águas superficiais. Portanto, apenas concentrações de ocorrência nessa matriz ambiental foram coletadas, ao passo que esta pesquisa excluiu experimentos de bancada realizados com águas residuárias sintética.

A construção da base de dados de ocorrência de fármacos contemplou registros de três classes terapêuticas: antibióticos, anti-inflamatórios e drogas psiquiátricas. Uma busca de dados

na base *Web of Science*® foi realizada para a coleta de artigos publicados durante todo o período de tempo disponível (1965 a 2021) considerando a ocorrência de fármacos em corpos hídricos em geral. Os artigos duplicados foram excluídos da amostra utilizando o software R. Posteriormente, foi realizada a leitura dos resumos para selecionar os artigos que consideravam a matriz de esgoto sanitário. Desses foram verificados os trabalhos que reportavam dados. Ao mesmo tempo, as referências citadas nos artigos selecionados também foram investigadas.

Ao se encontrar artigos com divulgação dos dados de monitoramento, as observações deveriam estar registradas em concentrações quantificáveis. Para a inclusão das observações na base, ao menos um dos seguintes critérios foi atendido: 1 – A concentração na entrada e a concentração na saída na ETE foi reportada, 2 – Quando a concentração na saída não foi reportada, obrigatoriamente, a concentração na entrada e a eficiência de remoção estavam explícitas no texto.

Quando as análises ocorreram em mais de uma etapa do tratamento (tratamento primário e secundário, por exemplo), registrou-se as concentrações e o tipo de técnica aplicada em cada uma das etapas específicas. A combinação de técnicas de tratamento em uma mesma etapa também foi registrada (associação de processos anaeróbios e aeróbios, por exemplo). Em geral, as concentrações foram reportadas por valores médios experimentais (média aritmética dos resultados quantificados nas análises experimentais).

4.2.5. Caracterização dos dados

A utilização de dados abertos é um dos fatores que favorece a transparência e a reprodutibilidade desta pesquisa. No entanto, é preciso considerar que os dados estão suscetíveis a imprecisões. Por isso, o Quadro 6 exhibe características das bases de dados que auxiliam a identificar possíveis fatores de incerteza neste trabalho.

Quadro 6 - Fatores de incerteza

	Anvisa	ANA	Dados da literatura científica
Coleta de dados	2021	2020	2021
Atualização da base no site	2023	2020	-
Utilização em pesquisas	Sim	Sim	Sim
Mesma localidade de interesse	Sim	Sim	Não
Procedência	Valores fornecidos por drogarias e farmácias	Visitas de campo	Experimentos reportados na literatura

Fonte: Elaborado pela autora.

A coleta de dados da ANVISA foi finalizada em 2021, quando a última planilha mensal relativa às comercializações de fármacos do mês de dez/2020 foi disponibilizada no site, que posteriormente, passou por novas atualizações com a finalidade de inserir os arquivos relativos ao ano de 2021. No entanto, a declaração dos registros passou a ser facultativa, em 2022, e deixou de ser obrigatória em 2023. Sendo assim, novas atualizações não foram verificadas.

Essa base de dados tem apresentado alta relevância para a realização de pesquisas, visto que os dados são disponibilizados de forma gratuita (CAETANO et al., 2022; CARRASCO; BARROS, 2024; LOPES; MOTTER; CARVALHO-SOARES, 2024). Outras bases comumente utilizadas como fonte de dados para trabalhos científicos são limitadas a instituições que realizam de acordos financeiros para compra.

Assim, a realização de pesquisas fica restrita a uma pequena parcela de pesquisadores que possuem os recursos para a aquisição de tais dados. Por isso, a base de dados do SNGPC (ANVISA) é uma ferramenta de democratização e ampliação do acesso a dados à uma ampla parcela da comunidade científica.

Em relação aos dados discutidos no Atlas Esgotos, o levantamento realizado por Lima et al. (2020) apontou que haviam sido utilizados em 129 publicações, até aquele período. No entanto, aproximadamente 70% das publicações consistia apenas na citação dos dados. É importante ressaltar que esses dados contêm informações específicas para as ETE de cada município do país.

Em 2020, foi disponibilizada uma atualização da base, resultado de solicitações de usuários do Atlas, instauração de novos Planos Municipais de Saneamento Básico, relatórios de fiscalização, registros de novos sistemas de saneamento, entre outras comunicações. Com isso, uma situação operacional atualizada do esgotamento sanitário foi disponibilizada para consulta e *download*.

Por fim, os dados coletados da literatura científica, provêm de experimentos realizados em diferentes localidades. É necessário considerar, que os procedimentos experimentais estão suscetíveis a incertezas relativas ao protocolo de amostragem ou erros humano e instrumentais.

Ademais, novos estudos podem surgir de localidades previamente não monitoradas, novos fármacos podem passar a ser identificados e, novos níveis de ocorrência podem ser determinados. Por isso, é preciso evidenciar que, eventualmente, uma nova coleta de dados da literatura deve ser realizada para representar de forma mais fidedigna a evolução da área.

4.3 IMPLEMENTAÇÃO DA MODELAGEM DE DADOS

Para a construção dos modelos, os valores anuais de PEC_{inf} ($ng \cdot L^{-1}$), unidades vendidas, quantidade por unidade (em unidades ou volume em mL), massa total comercializada (kg/ano), presentes na base de comercialização de fármacos, foram associados ao Fator de Excreção Humano (%), concentração de saída (em ngL^{-1}) e Eficiência de remoção (%) de cada fármaco considerado.

Propôs-se a construção de modelos para Carbamazepina e modelos para os antibióticos Amoxicilina, Claritromicina, Azitromicina, Cefalexina, Ampicilina, Ofloxacino, Norfloxacino e Ciprofloxacino, visto que são classes terapêuticas diferentes, com padrões de remoção diferentes.

A partir dos valores numéricos da concentração de saída gerou-se duas classes para a variável explicada “classe da concentração final de fármaco”: aceitável (1) ou preocupante (0). A predefinição das classes considerou os limites regulatórios para as concentrações de ocorrência dos fármacos considerados em água superficial de $100 ng \cdot L^{-1}$, $500 ng \cdot L^{-1}$ e $1000 ng \cdot L^{-1}$ (U. S. FOOD AND DRUG ADMINISTRATION, 1998; WIRTZ, 2009). Assim, cada limite regulatório é considerado um caso de investigação.

O método de partição de dados *Holdout* foi considerado e, portanto, aleatoriamente, foram selecionados os dados para os conjuntos de treinamento e teste na proporção 80:20. Para lidar com o potencial desbalanceamento de classes, a técnica de reamostragem *Synthetic Minority Oversampling Technique (SMOTE)* (CHAWLA *et al.*, 2002) foi selecionada para aplicação. O algoritmo da SMOTE é disponibilizado pela biblioteca *imbalanced-learn*, e sua proposta é gerar pontos de observações sintéticos com base nos valores dos vizinhos mais próximos (LEMAËŽTRE; NOGUEIRA; ARIDAS, 2017).

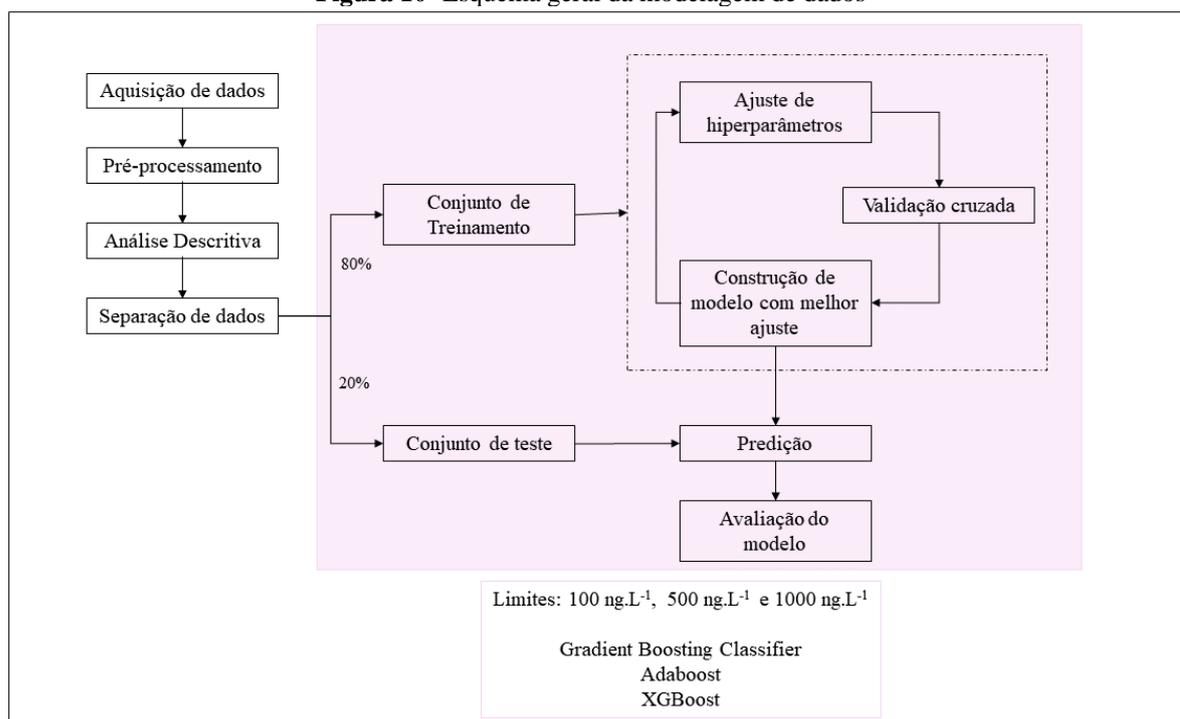
O método *GridSearchCV* foi utilizado para encontrar o estimador com o melhor ajuste de hiperparâmetros. Assim, foi passada uma faixa de valores de diferentes hiperparâmetros à função, que por meio de validação cruzada avaliou todas as combinações possíveis e, por fim, indicou o estimador que atendeu a essa condição. Após a identificação do estimador ótimo, a validação cruzada *K-fold*, com $k = 5$ partições, foi utilizada para avaliação do seu desempenho. Esse método particionou o conjunto de treinamento em 5 subconjuntos e, realizou o treinamento e avaliação dos modelos 5 vezes, cada vez selecionando 1 dos subconjuntos para validação. Após a avaliação, esse estimador foi considerado como modelo final para a predição, cujo desempenho foi avaliado junto ao conjunto de teste.

A linguagem *Python* foi escolhida para a construção dos modelos e a biblioteca *scikitlearn* para implementar os algoritmos Classificador Boosting Gradiente (*Gradient Boosting Classifier*) e Adaboost. A biblioteca XGBoost foi escolhida para implementar o algoritmo XGBoost. Esses algoritmos implementam as equações dos modelos descritos na sessão 3.3.2.

A função “*classification_report*”, da biblioteca *scikit-learn*, foi utilizada para verificação da Acurácia, Precisão, *Recall* e *F1 – score*, para cada classe em cada caso ajustado. O F1-score considerado para os modelos se refere ao “*macro averaged*”, que compreende a média aritmética do F1-score obtido para cada classe. Essa métrica, foi escolhida para garantir que ambas as classes apresentassem a mesma importância na avaliação. O método SHAP foi selecionado para auxiliar a interpretabilidade dos resultados e implementado pela biblioteca SHAP (*SHapley Additive exPlanations*). Por fim, foi realizada uma análise de sensibilidade.

As etapas realizadas para a construção do modelo podem ser visualizadas na Figura 10.

Figura 10- Esquema geral da modelagem de dados



Fonte: Elaborado pela autora.

4.3.1 Modelagem proposta para Carbamazepina

Os modelos propostos para a Carbamazepina consideraram como variáveis os valores anuais de PEC_{inf} (ng. L⁻¹), eficiência de remoção do tratamento da ETE e, a concentração de

saída (ng. L⁻¹). Como informado anteriormente, três casos foram desenvolvidos com o objetivo de se obter a classe da concentração final de fármaco a partir de diferentes limites normativos. Desse modo, o valor de concentração de saída é convertido em 0 ou 1 com base no valor de limite normativo considerado.

Para o primeiro caso, os valores de concentração de saída que se encontravam acima do limite de 100 ng. L⁻¹ recebiam a classe 0 e, observações com valor de concentração de saída igual ou menor que o valor considerado, recebiam a classe 1. O mesmo procedimento foi realizado para os demais casos, baseados na codificação de classes a partir dos limites de 500 ng. L⁻¹ e 1000 ng. L⁻¹.

Após a predefinição de classes para cada caso, o total de 434 observações foi dividido em conjunto de treinamento e teste, na proporção 80:20. Para os modelos baseados nos limites de corte 1000 ng. L⁻¹ e 100 ng. L⁻¹, a técnica SMOTE foi utilizada para o balanceamento de classes do conjunto de treinamento.

Assim, os três algoritmos de interesse foram executados para cada caso de limite de classificação e, ao final, as métricas de desempenho e o SHAP foram utilizados para avaliar e interpretar os resultados obtidos.

4.3.2 Modelagem proposta para os Antibióticos

Os modelos propostos para os Antibióticos consideram como variáveis os valores anuais de PEC_{inf} (ng. L⁻¹), unidades vendidas, quantidade por unidade (em unidades ou volume em mL), massa total comercializada (kg/ano), presentes na base de comercialização de fármacos, foram associados ao Fator de Excreção Humano (%), concentração de saída (em ngL⁻¹) e Eficiência de remoção (%).

Para a predefinição das classes, o mesmo procedimento adotado para a Carbamazepina foi implementado para os Antibióticos, com base nos mesmos limites normativos. Após a predefinição de classes, as 553 observações foram separadas em conjunto de treinamento e teste, na proporção 70:30. Para os modelos baseados nos limites de corte 1000 ng. L⁻¹ e 500 ng. L⁻¹, a técnica SMOTE foi utilizada para o balanceamento de classes e, posteriormente, os três algoritmos de interesse foram executados para os casos avaliados.

As métricas de desempenho e a biblioteca SHAP foram utilizadas para avaliar e interpretar os resultados obtidos.

4.3.3 Análise de Sensibilidade

Para ambos os casos descritos, em cada limite de classificação foi selecionado o modelo com melhor desempenho para realização de Análise de Sensibilidade, afim de verificar se variações nos valores anuais de PEC_{inf} (ng. L⁻¹) poderiam ocasionar alterações críticas nas classes de concentração final preditas.

Esta abordagem se justifica pelo fato de que a estimativa de PEC_{inf} (ng. L⁻¹) considera os valores de volume de esgotos gerado por dia (m³) catalogados pela ANA, como descrito no item 4.2.3. É importante destacar que é comum que as Estações de Tratamento de Esgotos apresentem ineficiências de projeto e de operação, o que pode acarretar em algum nível de incerteza nos dados divulgados.

Nesse sentido, é relevante desempenhar a análise de sensibilidade como forma de assegurar que os dados utilizados são válidos e, portanto, os resultados obtidos são confiáveis. Ainda, embora o Atlas Esgotos seja citado em pesquisas científicas, seus dados foram aplicados em apenas 4% das publicações investigadas (LIMA *et al.*, 2020).

Portanto, os modelos convencionais desenvolvidos nesta pesquisa consideraram os valores de volume de esgoto gerado por dia (m³) disponibilizados pela ANA para cada município. Para os modelos da Análise de Sensibilidade, considerou-se mais 3 níveis de variação: 25%, 50% e 75% do valor estimado pela ANA. Para cada nível de variação, o valor final de volume de esgotos foi calculado e, em seguida, novos valores de PEC_{inf} (ng. L⁻¹) foram estimados e, o procedimento de modelagem foi realizado novamente.

5 RESULTADOS E DISCUSSÃO

5.1 DADOS ANVISA – COMERCIALIZAÇÃO ANUAL DE FÁRMACOS

Os dados coletados da ANVISA foram “filtrados” para os municípios representativos, apresentando 24.729.021 observações. Ao se realizar o processo de exclusão das observações faltantes da coluna “PRINCIPIO_ATIVO”, o total de observações passou a ser 24.659.576. Posteriormente, os registros foram, novamente, “filtrados” para os fármacos de interesse e, ao final, foram obtidas 8.545.478 observações correspondentes à comercialização para o período de janeiro de 2014 à dezembro de 2020.

Os fármacos selecionados para esta investigação compreendem duas classes terapêuticas que estão entre as mais comercializadas mundialmente, os antibióticos e drogas psiquiátricas. Adicionalmente, a relevância dos fármacos considerados é legitimada pelo fato de que estão presentes em listas prioritárias de poluentes ou são apontados como potenciais candidatos à inclusão nessas listas.

Portanto, o alto padrão de comercialização desses fármacos, alta frequência de investigação na literatura científica, incapacidade de remoção completa no tratamento de ETE e os efeitos tóxicos que podem ocasionar ao serem lançados no corpo receptor, corroboraram a exploração dos antibióticos Amoxicilina, Ampicilina, Azitromicina, Cefalexina, Ciprofloxacino, Claritromicina, Norfloxacino e Ofloxacino, bem como a droga psiquiátrica Carbamazepina.

Os dados apontaram maior taxa de comercialização para os fármacos na forma sólida apresentados em caixas, com exceção do Ofloxacino que é apresentado como solução oftalmológica (embora tenham sido constatados registros que indiquem “caixa”, equivocadamente). Assim, para cada fármaco, foi possível identificar a quantidade de unidades e a forma de apresentação mais comercializada para a população entre 2014 e 2020, como indicado na Tabela 5.

Tabela 5 – Dispensação de fármacos por forma de apresentação, acumulada para os anos de 2014 a 2020.

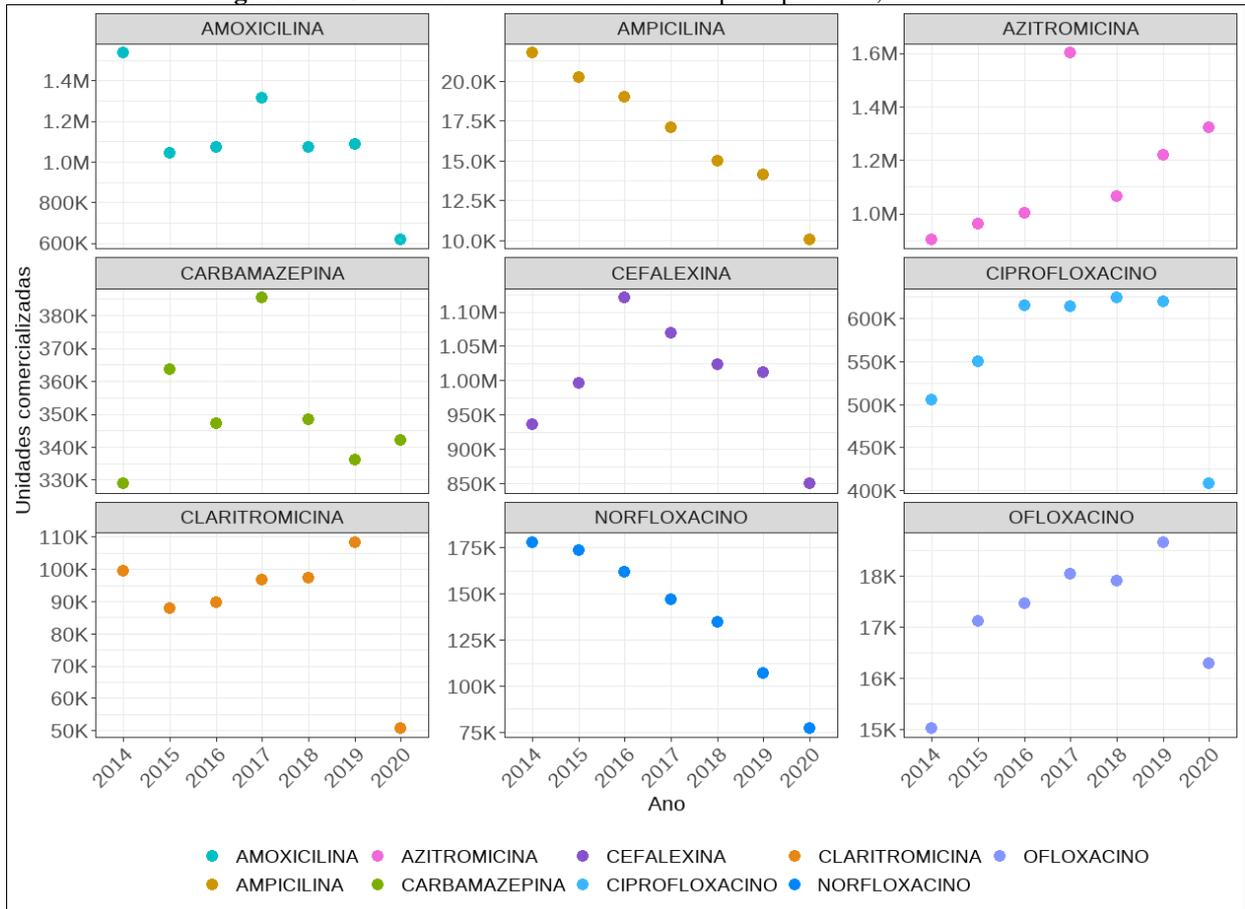
Fármaco	Caixas (unidades)	Caixas (%)	Frascos (unidades)	Frascos (%)
Amoxicilina	6.595.323	85,19	1.146.665	14,81
Ampicilina	111.008	94,71	6.205	5,29
Azitromicina	7.450.169	92,22	628.267	7,78
Carbamazepina	2.287.919	93,32	163.759	6,68
Cefalexina	6.511.965	92,93	495.428	7,07
Ciprofloxacino	3.746.527	95,18	189.858	4,82
Claritromicina	562.237	89,24	67.803	10,76

Fármaco	Caixas (unidades)	Caixas (%)	Frascos (unidades)	Frascos (%)
Norfloxacino	918.646	93,82	60.537	6,18
Ofloxacino	-	-	120528	100

Fonte: Elaborado pela autora a partir de dados da ANVISA.

Na Figura 11, mostra-se a tendência de comercialização de unidades em frascos e caixas de todos os fármacos para o período considerado.

Figura 11 - Unidades comercializadas de cada princípio ativo, anualmente.



Fonte: Elaborado pela autora a partir de dados da ANVISA.

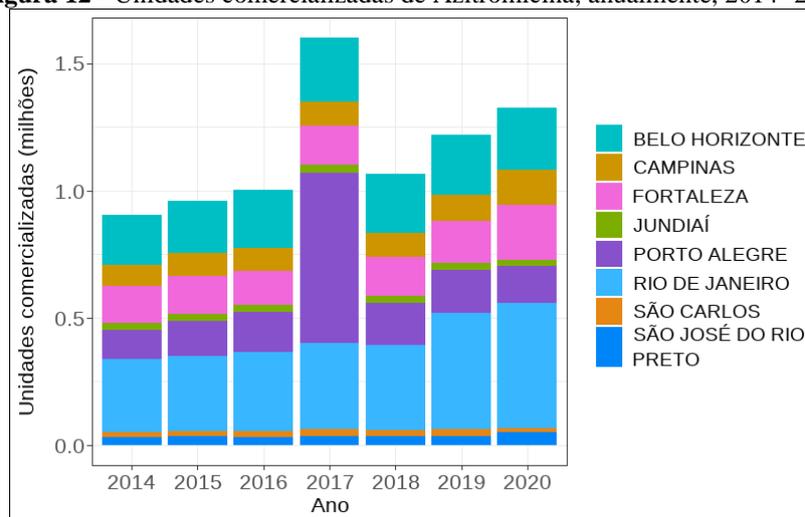
É possível observar que a comercialização de alguns fármacos apresentou tendência de aumento ao longo do tempo, como Azitromicina, Claritromicina, Ofloxacino e Ciprofloxacino. Esses três últimos fármacos também apresentaram uma queda nas unidades comercializadas no ano de 2020. A Ampicilina e o Norfloxacino apresentaram queda nos índices de comercialização. A Amoxicilina e Cefalexina também apresentaram queda do índice de comercialização no ano de 2020.

Com exceção da capital Fortaleza, da região Nordeste, e Porto Alegre, da região Sul, os municípios selecionados para a investigação pertencem ao Sudeste. Segundo Lopes *et al.*

(2024), mais da metade da comercialização de antimicrobianos para o período de 2014 – 2019 pertence a essa região, ainda que tenha sido verificado aumento geral de 30% na comercialização em todas as regiões do Brasil, associado à densidade populacional e renda.

O antibiótico Azitromicina apresentou o total de 8.078.436 unidades vendidas para o período investigado. Os resultados apresentados na Figura 12 mostram o padrão anual de comercialização da Azitromicina por município investigado e apontam um pico de comercialização em 2017, devido ao brusco aumento de vendas na cidade de Porto Alegre. Entre 2018 e 2020 o padrão de vendas em Porto Alegre se reduziu aos índices comumente identificados, mas, verificou-se um aumento gradual na quantidade comercializada. A cidade do Rio de Janeiro compreendeu 31,2% das unidades comercializadas de Azitromicina, e Belo Horizonte 19,2% da comercialização.

Figura 12 - Unidades comercializadas de Azitromicina, anualmente, 2014 -2020



Fonte: Fonte: Elaborado pela autora a partir de dados da ANVISA.

Caetano *et al.* (2022) verificaram que o grupo dos macrolídeos, que compreende os fármacos investigados Azitromicina e Claritromicina, foi o segundo mais comercializado entre 2017 e 2019, no Brasil. Em 2020, esse grupo apresentou aumento considerável nas vendas, devido ao início da pandemia, e a Azitromicina revelou-se o fármaco mais comercializado (24,6% do total).

Com base nas doses médias de ingestão calculadas, os autores apontaram um consumo exorbitante desse fármaco pela população. O mesmo padrão foi observado por Trindade *et al.* (2023) para a cidade de Aracaju (SE), onde a comercialização de Amoxicilina foi reduzida entre 2019 e 2020, enquanto a de Azitromicina aumentou em 52% no mesmo período, passando a liderar as vendas.

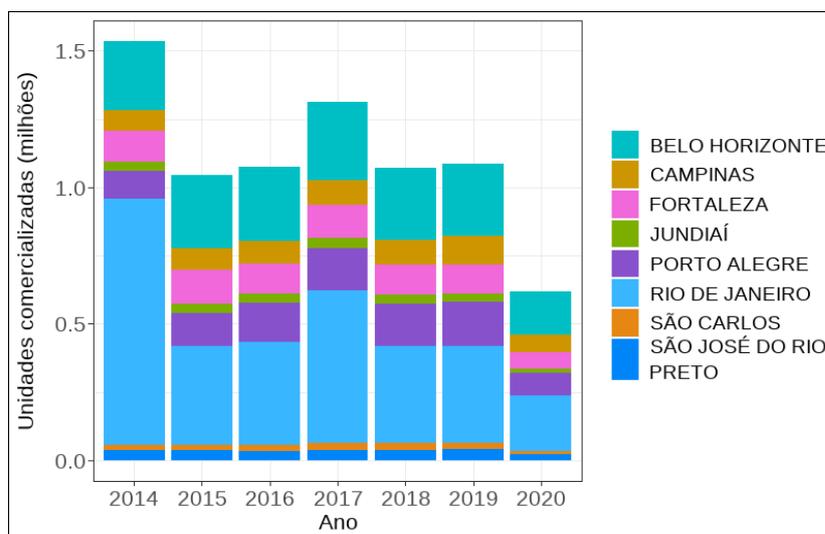
Esses resultados concordam com a tendência observada para a Azitromicina, nessa pesquisa. Entretanto, o pico de comercialização identificado no ano de 2017 para a cidade de Porto Alegre, aponta a dispensação exacerbada do fármaco, em quantidades que superam o tratamento individual. Não foram encontrados trabalhos específicos em nível estadual ou municipal que esclarecessem o perfil observado. No entanto, a alta comercialização de fármacos na região Sul é citada na literatura e, comumente, é associada ao poder econômico da população.

Lima *et al.* (2017b) verificaram que para dados referentes ao período de Julho de 2014 a Maio de 2015, a região Sul apresentou a maior média de medicamentos prescritos (2,9), superando a média nacional (2,4). Lopes *et al.* (2024) identificaram que entre os anos 2014 e 2020, os dados do SNGPC apontaram a região Sul como líder do consumo de antimicrobianos baseado em doses diárias definidas (*Defined daily doses per 1000 inhabitants per day, DDI*s), e em segundo lugar em relação ao volume de vendas. Em ambos os casos, o perfil observado foi associado à alta densidade populacional e renda.

Desse modo, a região Sul apresenta destaque pela alta contribuição ao índice de comercialização de fármacos, o que confirma a necessidade de investigações aprofundadas dos dados do SGNPC que ponderem as ocorrências da região para os anos investigados.

A Amoxicilina foi o segundo fármaco mais comercializado no período, apresentando um total de 7.741.988 unidades vendidas. A maior taxa de comercialização foi identificada no ano de 2014, devido a ampla comercialização no município do Rio de Janeiro, como pode ser verificado na Figura 13.

Figura 13 - Unidades comercializadas de Amoxicilina anualmente, 2014 - 2020.



Fonte: Elaborado pela autora a partir de dados da ANVISA.

Exceto pelo ano de 2014 e 2017, é possível verificar que as vendas se mantiveram estáveis nesse município, sofrendo uma redução no ano de 2020. Ainda assim, nota-se que o Rio de Janeiro foi responsável por 40,2% das unidades comercializadas de Amoxicilina. Seguidamente, Belo Horizonte representou 22,3% da comercialização.

A tendência observada para a Amoxicilina condiz com Caetano *et al.* (2022), que apontaram o aumento da comercialização da classe das penicilinas entre 2017 e 2019, especialmente aquelas que apresentavam o inibidor beta-lactamase, seguido por significativa redução no ano de 2020, início da pandemia.

Observou-se pico de comercialização nos anos de 2014 e 2017, com a cidade do Rio de Janeiro sendo a que apresentou maiores índices de venda, como exibido na Figura 13. Nesse caso, um dos fatores que pode ter influenciado é o fato de que o Rio de Janeiro é a cidade investigada de maior população. Além disso, o estado possui sua relevância no mercado farmacêutico evidenciada pela alta contribuição no faturamento total, que para o ano de 2017 correspondeu a 11,21%, superado apenas pelo estado de São Paulo (ANVISA, 2018).

No entanto, fatores como prescrições indevidas e a extrapolação de quantidades individuais também devem ser considerados. O Relatório Anual Proteste (ASSOCIAÇÃO DOS CONSUMIDORES, 2015) reportou a prescrição de antibióticos, por parte de médicos, a pacientes saudáveis, na cidade do Rio de Janeiro. Esse comportamento evidencia o uso de práticas inadequadas na prescrição de antibióticos, discutido amplamente na literatura (LIMA *et al.*, 2017b; TRINDADE *et al.*, 2023; VILLALOBOS *et al.*, 2024)

Adicionalmente, para dados de julho de 2014 a maio de 2015, Lima *et al.* (2017) identificaram a dispensação de prescrições inadequadas de antibióticos em Unidades Básicas de Saúde (UBS), em diversas localidades, inclusive Rio de Janeiro, e citaram que 74,8% dos pacientes investigados reportaram falta de instrução em relação a administração dos antibióticos prescritos.

O alto índice de comercialização da Amoxicilina também foi identificado por Moura *et al.* (2022), que observou um valor aproximadamente 3 vezes maior do que outros antibióticos investigados, considerando dispensações no período de 2008 a 2016 para a região metropolitana de São Paulo. No entanto, foi verificada a redução na comercialização desse fármaco, associada a implementação da RDC 44 de 26 de Outubro de 2010 (ANVISA, 2010), que estabeleceu a comercialização de antimicrobianos diante de prescrição médica.

Assim, a ampla comercialização dos fármacos Azitromicina e Amoxicilina pode ser justificada pelo fato de serem de amplo espectro, ou seja, possuem ação inibitória a uma variedade de bactérias. A Azitromicina é prescrita para tratamento de infecções respiratórias e

pneumonias (MOURA *et al.*, 2022), apresenta efeito imunomodulador e anti-inflamatório (PANI *et al.*, 2020). A Amoxicilina é administrada especialmente para tratamentos de diferentes infecções como de garganta, nariz, ouvido, respiratórias e do trato urinário, pele e outros usos.

Além da Azitromicina e Amoxicilina, a Cefalexina e Ciprofloxacino também apresentaram um perfil relevante de vendas em comparação aos outros fármacos selecionados. Esse resultado concorda com o verificado por Lopes *et al.* (2024), que investigaram os dados do SNGPC para todas as regiões brasileiras e apontaram que esses quatro compostos foram os mais comercializados para os anos de 2014 a 2019. Neves e Castro *et al.* (2020) também verificaram a Amoxicilina, Cefalexina e Azitromicina como os fármacos mais comercializados para o período de 2013 a 2016, em todo o Brasil.

Nesse sentido, o Ciprofloxacino é um fármaco da subclasse das fluoroquinolonas, amplamente utilizado para o tratamento de infecções renais e do trato urinário (VILLALOBOS *et al.*, 2024), doenças sexualmente transmissíveis, infecções respiratórias e, ainda, infecções dentárias, embora nesse caso a prescrição seja questionada entre os profissionais da área e, a depender do caso, considerada indevida (CARRASCO; BARROS, 2024).

Moura *et al.* (2022) verificaram o aumento de vendas do Ciprofloxacino para o período de 2008 a 2016, na região metropolitana de São Paulo. Esse resultado concorda com a tendência verificada na Figura 11, onde foi observado um aumento notável de unidades comercializadas de 2014 a 2016. Carrasco e Barros (2024) também identificaram ampla comercialização para o Ciprofloxacino no período de 2015 a 2021, ocupando o 10º lugar dentre os antibióticos mais comercializados no Brasil, considerando suas formas combinadas.

A Cefalexina é utilizada amplamente no tratamento de infecções respiratórias. Esse fármaco foi identificado, no estudo Trindade *et al.* (2023), como o segundo antibiótico de maior comercialização na cidade de Aracaju, entre 2014 e 2020, superando 130 mil unidades comercializadas. Outros autores também apontaram a alta incidência de vendas desse fármaco, posicionando-o entre os antibióticos com maior índice de comercialização (LOPES; MOTTER; CARVALHO-SOARES, 2024; MOURA *et al.*, 2022; NEVES E CASTRO *et al.*, 2020).

Em síntese, com relação a quantidade de unidades dispensadas de 2014 a 2020 verificada nessa pesquisa, o fármaco Azitromicina correspondeu a 26,0% do valor total, seguido por Amoxicilina (24,92%), Cefalexina (22,56%), Ciprofloxacino (12,67%), Carbamazepina (7,89%), Norfloxacino (3,15%), Claritromicina (2,02%), Ofloxacino (0,38%) e Ampicilina (0,37%).

Foi demonstrado pelos dados investigados que, com exceção da Azitromicina e Carbamazepina, os fármacos apresentaram queda na comercialização no ano de 2020, início da

pandemia. Caetano *et al.* (2022) identificaram redução de 32,7% na comercialização de antibióticos no ano de 2020, primordialmente no Sul e Sudeste, regiões que contemplam a maior parte dos municípios investigados na presente pesquisa. Um dos fatores de influência apontado pelos autores foi a crise econômica do país associada à obrigatoriedade de apresentação de prescrição para a compra.

A Carbamazepina compreendeu 7,89% do total de unidades comercializadas entre 2014 e 2020. Esse fármaco participa da classe terapêutica dos medicamentos psicotrópicos e subclasse de anticonvulsivantes, que atuam no sistema nervoso central para regular alterações de humor, emoções, comportamento e controle de crises convulsivas (MORENO *et al.*, 2004). O crescimento do consumo dessa classe terapêutica tem sido apontado com preocupação diante dos efeitos colaterais que podem ocasionar ao ser humano, inclusive a dependência química (DE OLIVEIRA SANTOS; BUENO; DE PASSOS, 2024; DEL FIOLE *et al.*, 2022; HOEFLER *et al.*, 2023).

Assini e Back (2017) avaliaram as dispensações de psicotrópicos no ano de 2014, em um município do interior do Estado de Santa Catarina e verificaram alto consumo por parte da população. A dispensação de anticonvulsivantes correspondeu a 12,2% do total, sendo a Carbamazepina responsável por 3,34% das dispensações totais.

No estudo de Silva *et al.* (2022), a Carbamazepina correspondeu a 15,5% das dispensações totais de psicotrópicos na cidade de Vitória da Conquista (BA), para o período de 2018, 2020 e 2021. Para cada ano avaliado, a Carbamazepina ocupou o terceiro lugar em quantidade de dispensações. Não foram encontrados estudos específicos sobre a dispensação de Carbamazepina para os municípios investigados nessa pesquisa, o que indica a necessidade de mais investigações a partir dos dados do SNGPC.

É importante ressaltar que esses fármacos são prescritos sob diferentes formas de apresentação, como comprimidos ou em solução. Nesse sentido, a forma de apresentação mais comercializada de cada fármaco é indicada na Tabela 6.

Tabela 6 - Forma mais comercializada de cada fármaco e o percentual equivalente nas vendas totais

Fármaco	Massa/ concentração	Unidades comercializadas (%)
Amoxicilina	500 mg	11,83
Ampicilina	500 mg	0,35
Azitromicina	500 mg	21,21
Carbamazepina	200 mg	6,05
Cefalexina	500 mg	19,17
Ciprofloxacino	500 mg	11,95
Claritromicina	500 mg	1,26
Norfloxacino	400 mg	3,15
Ofloxacino	3 mg/ml	0,38

Fonte: Elaborado pela autora a partir de dados da ANVISA.

Verifica-se que com exceção do Ofloxacino, que é comercializado em solução, a forma mais comercializada de todos os outros fármacos é em comprimido. Além disso, para a maioria dos antibióticos investigados, a forma mais comercializada contém 500 mg do composto, como é o caso da Amoxicilina, Ampicilina, Azitromicina, Cefalexina e Claritromicina. A forma mais comercializada do Norfloxacino contém 400 mg de fármaco em cada comprimido. Em conjunto, as formas mais comercializadas correspondem a aproximadamente 75% do total de unidades comercializadas.

Nesse sentido, o percentual restante de unidades, é comercializado sob diferentes apresentações, a depender do fármaco. Foram identificadas apresentações em comprimido e solução para Amoxicilina, Ampicilina, Azitromicina, Cefalexina, Ciprofloxacino e Carbamazepina, o que totaliza em 16, 4, 10, 9, 10 e 3 formas distintas de apresentação, respectivamente. Foi identificada apenas uma forma de comercialização do Norfloxacino, em comprimidos e, uma forma em solução para o Ofloxacino.

Resultados semelhantes foram apontados por Lopes *et al.* (2024), que verificaram diferentes formas de apresentação da Amoxicilina, Ciprofloxacino, Azitromicina, Cefalexina e Ampicilina, sendo essa ordem da maior quantidade para a menor quantidade de apresentações.

Cada forma de apresentação indica a quantidade de comprimidos na caixa ou o volume total da solução presente no frasco. Essa informação é relevante, pois, essa quantidade impacta diretamente na massa consumida de fármaco. Em complementação, o número de unidades de embalagens comercializadas (caixas ou frascos) também é relevante para estimar a massa consumida. Assim, os valores correspondentes da massa comercializada são resumidos na Tabela 7.

Tabela 7 - Massa de fármacos (kg) comercializada por ano nos municípios representativos, 2014 - 2020

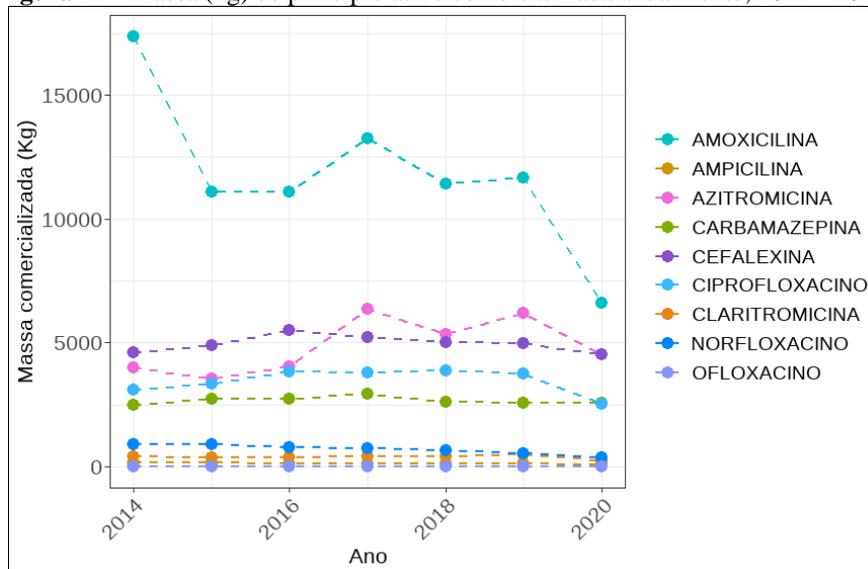
Fármaco	Total	Mínimo	Máximo	Média	Desvio- padrão
Carbamazepina	18.610,08	2.470,70	2.924,12	2.658,58	140,67
Ciprofloxacino	24.235,80	2.525,34	3.894,06	3.462,26	468,19
Amoxicilina	82.432,28	6.591,63	17.363,09	11.776,04	2.961,55
Ofloxacino	1,81	0,225	0,280	0,26	0,02
Claritromicina	2.697,35	260,25	477,46	385,33	61,43
Cefalexina	34.812,74	4.541,57	5.216,32	4.973,25	311,37
Norfloxacino	4.866,42	381,32	895,99	695,20	174,89
Ampicilina	900,05	80,01	164,43	128,58	26,62
Azitromicina	33.973,94	6.368,31	3.535,05	4.853,42	1.039,08

Fonte: Elaborado pela autora a partir de dados da ANVISA.

A Amoxicilina se destaca por apresentar a maior massa comercializada, o que corrobora com a alta dispensação de unidades verificada. Em contrapartida, o Ofloxacino foi o fármaco

com o menor índice de comercialização e, conseqüentemente, menor massa acumulada (1,81 kg). A variação anual de massa para cada fármaco é apresentada na Figura 14.

Figura 14 - Massa (kg) de princípio ativo comercializada anualmente, 2014 - 2020.



Fonte: Elaborado pela autora a partir de dados da ANVISA.

O padrão observado para a Amoxicilina confirma que a sua comercialização supera a dos outros fármacos e, é notória a redução de massa devido à queda da comercialização no ano de 2020. A massa comercializada de Azitromicina apresenta uma tendência geral de aumento. Porém, mesmo sendo o fármaco com maior quantidade de unidades dispensadas em 2020, em termos de massa comercializada esse fármaco ainda foi superado pela Amoxicilina. Esse resultado indica a influência da variação de doses e quantidade de comprimidos contidos nas unidades dispensadas.

Os fármacos Ofloxacino, Norfloxacino, Claritromicina e Ampicilina, assim como apresentaram menores índices de comercialização, demonstraram valores reduzidos de massa comercializada, mantendo-se abaixo de 1000 kg em todos os anos.

A Tabela 7 apresentada, mostra os valores agrupados para todos os municípios e todos os anos. No entanto, os cálculos foram realizados em base anual para cada município separadamente, a fim de obter uma estimativa de PEC_{inf} acurada. Nesta pesquisa, considerou-se que as unidades dispensadas foram completamente consumidas, embora essa não seja a única variável relevante para a disponibilização desses compostos quando excretados para a rede coletora de esgotos.

É importante salientar que a avaliação da comercialização dos fármacos à luz do período pandêmico não contemplou os objetivos dessa pesquisa. No entanto, a consideração dos dados

referentes ao ano de 2020, fez necessário pontuar os padrões atípicos observados. Notou-se, primordialmente, o aumento significativo das comercializações da Azitromicina, que recomendada em conjunto com hidroxicloroquina, ivermectina, nitazoxanida e outros, integrou o Kit-Covid, lamentavelmente disseminado por autoridades brasileiras como tratamento precoce ao contágio pelo vírus SARS-CoV-2, mesmo sem evidências científicas.

5.2 CONCENTRAÇÃO AFLUENTE (PEC_{INF}) PARA OS MUNICÍPIOS

A partir da segregação da forma de apresentação, calculou-se a massa anual de fármacos comercializados e, posteriormente, a parcela PEC_{inf} , correspondente à concentração afluente de cada fármaco encaminhada nos esgotos para tratamento nas ETE. A consideração dessa variável é importante, já que a mesma é expressa a partir do perfil de comercialização dos fármacos. Para o cálculo, considerou-se também a população, vazão de esgoto afluente e o fator de excreção humano. Indica-se na Tabela 8 a população, vazão afluente total e a vazão per capita de cada município.

Tabela 8 - População e vazão afluente de esgoto sanitário.

Município	População	Vazão afluente do município ($L.s^{-1}$)	Vazão per capita ($m^3. (dia.hab)^{-1}$)
Jundiá	376.974	915,7	0,209
São José do Rio Preto	407.707	1.050	0,222
Rio de Janeiro	6.429.923	9.461,3	0,127
Belo Horizonte	2.479.165	3.752,5	0,130
Fortaleza	2.551.806	2.591	0,088
São Carlos	226.987	550	0,209
Porto Alegre	1.467.816	2.101	0,124
Campinas	1.125.176	1.391,7	0,107

Fonte: Atlas Esgoto (ANA, 2017).

Segundo informações dos registros da ANA, a vazão afluente para a situação de coleta e tratamento de esgotos considera todas as ETE catalogadas para cada município. A unidade “ $m^3. (dia.hab)^{-1}$ ” foi utilizada, pois a PEC_{inf} foi expressa em $ng. L^{-1}$. Nesse sentido, observa-se na Tabela 9 os valores anuais totais de PEC_{inf} , correspondentes ao somatório em bases anuais para cada município considerado nessa pesquisa.

Tabela 9 – Concentração anual afluente de fármaco – PEC_{inf} ($ng. L^{-1}$)

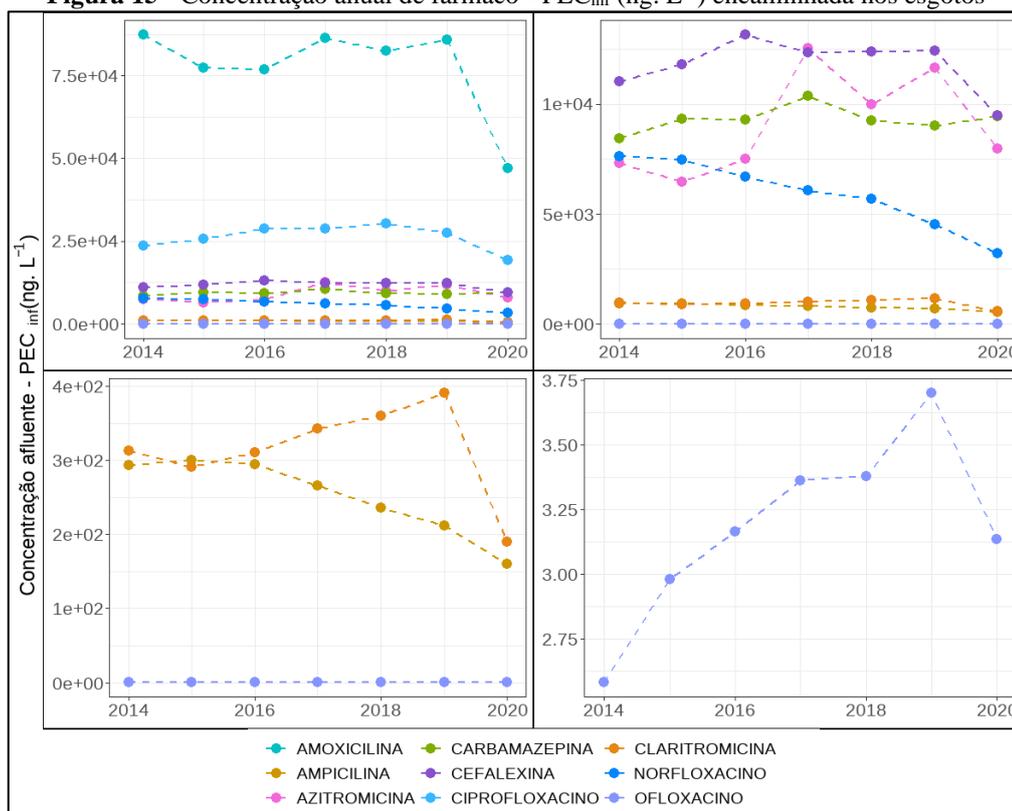
Fármaco	2014	2015	2016	2017	2018	2019	2020
Carbamazepina	8.428,66	9335,25	9.287,90	10.366,64	9.229,34	9.015,20	9.431,39
Ciprofloxacino	23.735,02	25.710,02	28.694,11	28.640,51	30.385,57	27.428,83	19.341,49
Amoxicilina	87.508,73	77.407,14	76.827,81	86.335,43	82.638,17	85.844,92	47.034,39
Ofloxacino	2,58	2,98	3,17	3,36	3,38	3,70	3,13
Clarithromicina	957,35	879,84	931,42	1.013,43	1.076,33	1.168,86	564,38

Cefalexina	11.025,67	11.797,74	13.171,06	12.357,31	12.400,07	12.409,60	9.467,55
Norfloxacino	7.611,55	7.480,19	6.674,33	6.081,28	5.679,72	4.541,59	3.208,22
Ampicilina	937,14	906,79	851,47	812,36	746,89	706,91	529,53
Azitromicina	7.314,33	6.475,07	7.507,30	12.525,85	9.988,81	11.659,51	7.983,26

Fonte: Elaborado pela autora.

Na Figura 15 é possível observar a tendência anual da concentração encaminhada ao tratamento em ETE.

Figura 15 - Concentração anual de fármaco - PEC_{inf} (ng. L⁻¹) encaminhada nos esgotos



Fonte: Elaborado pela autora.

Em concordância ao perfil de quantidade de unidades e massa comercializada, a Amoxicilina apresentou a maior concentração anual encaminhada junto aos esgotos para tratamento em ETE, seguida pelo Ciprofloxacino. A Claritromicina, Ampicilina e Ofloxacino apresentaram baixas concentrações, como consequência da reduzida comercialização em detrimento aos outros fármacos.

Portanto, os valores de concentração apresentados são decorrentes da coleta e manipulação dos dados da ANVISA, ANA e, literatura científica (Fator de Excreção). Esses valores são de particular interesse para a consideração nos modelos construídos.

5.3 DADOS DA LITERATURA – CONCENTRAÇÕES DE OCORRÊNCIA EM ESGOTOS SANITÁRIOS

Um total de 64 trabalhos que analisaram 55 fármacos em 23 países foram compilados na base de dados de concentrações de ocorrência de fármacos provenientes de experimentos publicados na literatura. Coletaram-se observações das classes terapêuticas de anti-inflamatórios, antibióticos e drogas psiquiátricas.

Contemplam-se registros de Autoria, Classe Terapêutica, Princípio Ativo, Local, Tecnologia de tratamento, concentração de entrada experimental (ng. L⁻¹), concentração de saída experimental (ng. L⁻¹) e Eficiência de remoção (%) no tratamento.

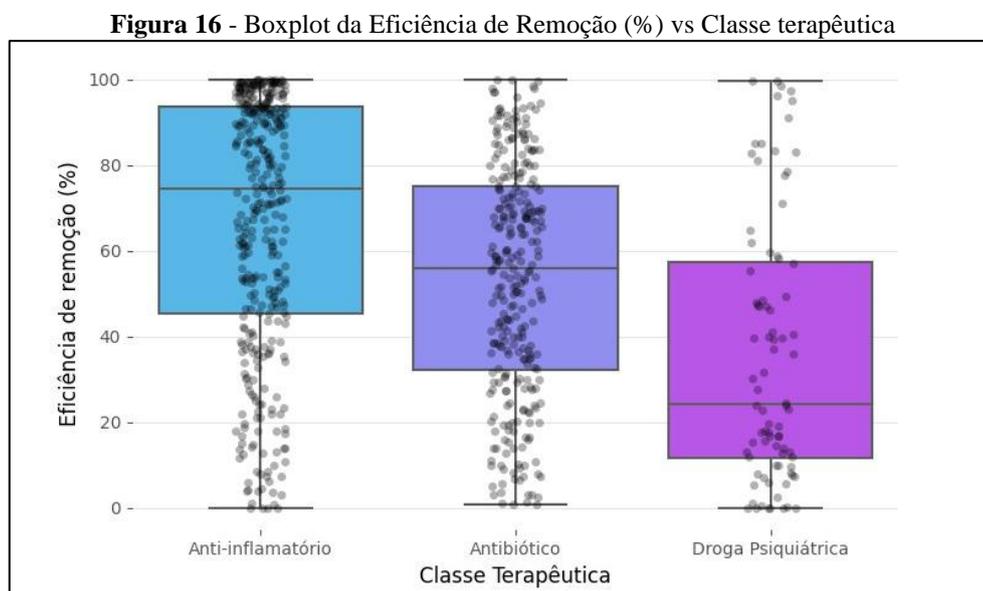
Após ajustes, conversões de unidades, formatação e uniformização de termos, a base de dados apresentou 941 observações. Verificaram-se um total de 417 observações relativas a anti-inflamatórios, 356 a antibióticos e 167 de drogas psiquiátricas. Os fármacos Carbamazepina, Diclofenaco, Naproxeno, Ibuprofeno e Sulfametoxazol foram identificados com maior recorrência na base, correspondendo a 58,5% das observações da base.

A China foi o país que apresentou maior número de observações, um total de 157. Em seguida, o Japão, Canadá e Brasil apresentaram, 109, 92 e 74 observações, respectivamente. A maior incidência de registros se deu para a técnica de tratamento de esgotos em nível secundário Lodo Ativado, recorrente em 413 observações e, seguidamente, para o tratamento *Submerged Attached Growth Reactor (SAGR, Reatores com biomassa aderida)*, com 56 observações provenientes da investigação de Anderson *et al.* (2020).

A classe de anti-inflamatórios apresentou maiores valores de concentração afluente em ETE, com o valor médio de 6.845,08 ng. L⁻¹, enquanto os antibióticos e drogas psiquiátricas apresentaram valores médios de concentrações afluentes de 1.160,95 ng. L⁻¹ e 1.472,66 ng. L⁻¹, respectivamente. No entanto, foi verificado para a concentração efluente de anti-inflamatórios o valor médio de 938,39 ng. L⁻¹, e para os antibióticos a concentração média de 928,69 ng. L⁻¹. A concentração média efluente de drogas psiquiátricas foi de 615,4 ng. L⁻¹, após o tratamento de remoção.

Foi verificado o acúmulo de poluentes para 198 observações. Entretanto, essas observações não foram utilizadas posteriormente, uma vez que indicam a possível formação de metabólitos ou compostos recalcitrantes. Em ambos os casos, não faz parte do escopo desta pesquisa a investigação desse aspecto.

Desse modo, desconsiderando-se as observações que apresentaram acúmulo de fármacos, os valores de eficiência de remoção observados para as classes terapêuticas consideradas, variaram como indicado na Figura 16.



Fonte: Elaborado pela autora a partir de dados coletados na literatura.

É possível notar que há uma considerável variabilidade nos valores de eficiência de remoção, uma vez que os dados refletem as condições experimentais empregadas pelos autores.

Verifica-se que os menores percentuais de eficiência de remoção se dão para a classe de drogas psiquiátricas, para a qual metade dos registros indicam remoção de apenas 24% da concentração inicial do composto.

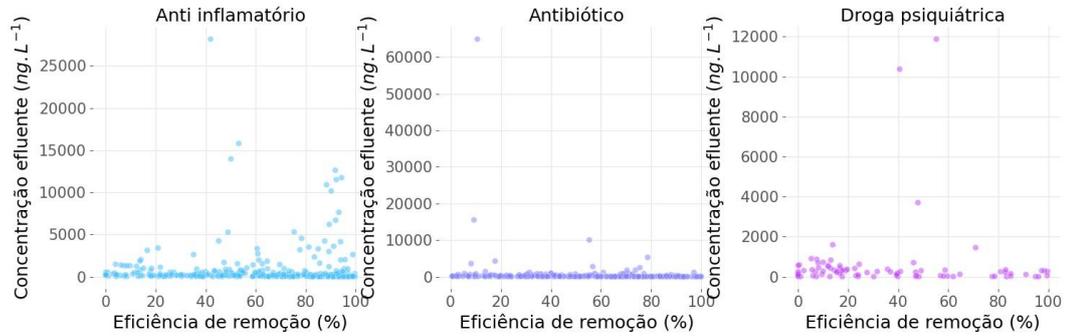
Esse fato indica que, dentre os relatos experimentais coletados, essa classe demonstrou ser a mais resistente ao tratamento biológico. Já a classe dos anti-inflamatórios apresentou tendência de maior eficiência de remoção, dada a aglomeração de pontos na região do terceiro quartil e segundo quartil, de valor 74%. Os antibióticos apresentaram um perfil de remoção intermediário, com mediana de 55%.

Na Figura 17 exibe-se a Eficiência de remoção dos fármacos pela Concentração efluente na ETE por classe terapêutica. A classe de antibióticos contém a observação com maior concentração efluente, equivalente a $64.900 \text{ ng. L}^{-1}$, valor verificado experimentalmente no México. As classes de anti-inflamatórios e drogas psiquiátricas também apresentaram altos valores de concentração efluente, com valor máximo de $28.147 \text{ ng. L}^{-1}$, verificado no Reino Unido, e $11.870 \text{ ng. L}^{-1}$, na Espanha, respectivamente.

Ainda, nota-se que para os anti-inflamatórios, há uma ocorrência de observações com alto valor de concentração efluente, mas que apresentaram eficiência de remoção de 80% a

100%. O mesmo não é observado para os antibióticos, cuja eficiência de remoção se apresenta bastante dispersa. Para as drogas psiquiátricas, é possível observar a incidência de observações com eficiência de remoção entre 0 a 20%.

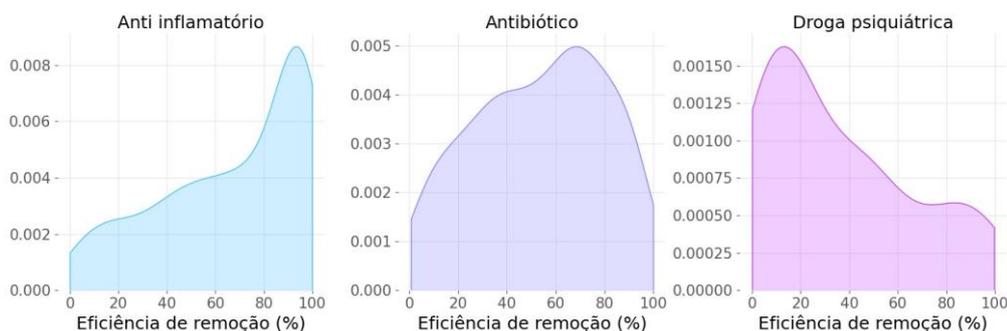
Figura 17 - Dispersão entre Eficiência de Remoção (%) e Concentração efluente por Classe terapêutica



Fonte: Elaborado pela autora a partir de dados coletados na literatura.

A curva de densidade da Eficiência de remoção para cada classe terapêutica é exibida na Figura 18, que corrobora com o comportamento identificado nas figuras anteriores. Em especial, a classe de antibióticos apresenta o perfil mais disperso entre todas. Porém, é possível identificar que entre 40% a cerca de 80% de eficiência de remoção, há uma região com maior densidade de probabilidade.

Portanto, a Eficiência de remoção por classe terapêutica aponta que as drogas psiquiátricas apresentam maior resistência ao tratamento, os antibióticos apresentam resistência intermediária e, os anti-inflamatórios são os fármacos com menor resistência ao tratamento, no caso avaliado.

Figura 18 - Gráfico de Densidade da Eficiência de Remoção (%) por Classe terapêutica

Fonte: Elaborado pela autora a partir de dados coletados na literatura.

Nesse sentido, México, Reino Unido e Espanha apresentaram as maiores concentrações máximas efluentes. Resultados oriundos do Quênia também mostraram altas concentrações, na faixa de 3.890 ng. L^{-1} – $39.340 \text{ ng. L}^{-1}$ para concentrações afluentes e, suas concentrações efluentes variaram na faixa de 1.170 ng. L^{-1} – $10.040 \text{ ng. L}^{-1}$. Juntamente com o México, apresentaram as maiores médias de concentração efluente: $4.072,86 \text{ ng. L}^{-1}$ (Quênia) e $5.113,32 \text{ ng. L}^{-1}$ (México). Dentre todos os dados coletados da literatura, a concentração de saída (ng. L^{-1}) e a eficiência de remoção (%) dos fármacos selecionados previamente, são de interesse para compor os modelos de classificação propostos.

5.4 IMPLEMENTAÇÃO DOS MODELOS

Duas abordagens de modelos foram desenvolvidas: para Carbamazepina e para os Antibióticos. Para cada uma, os algoritmos *Gradient Boosting Classifier*, *Adaboost Classifier* e *XGBoost* foram implementados para classificação em três limites de corte para rotulação de classes: 100 ng. L^{-1} , 500 ng. L^{-1} e 1.000 ng. L^{-1} .

Uma vez que foi feita uma associação entre os valores de comercialização de fármacos de cada ano provenientes da ANVISA e as concentrações de ocorrência coletadas na literatura, a quantidade final de observações para a Carbamazepina foi 434 e de antibióticos foi 553. Além disso, foi observado desbalanceamento ao gerar as classes de concentração final para os limites regulatórios considerados. O desbalanceamento faz com que o modelo tenda a classificar as amostras de teste apenas como a classe majoritária, revelando resultados enviesados.

Desse modo, os ajustes incluíram a etapa de balanceamento das classes no conjunto de treinamento, pelo método SMOTE. Por fim, a construção dos modelos contou com as etapas de implementação dos algoritmos, ajuste de hiperparâmetros, avaliação de desempenho pelas

métricas Precisão, *Recall*, *F1-score* e Acurácia. Adicionalmente, a influência do pré-tratamento de dados também foi avaliada e o método de melhor desempenho foi identificado.

5.4.1 Modelos para a Carbamazepina

5.4.1.1 Métricas de Avaliação

Os algoritmos selecionados para este estudo foram Modelo Boosting Gradiente (*Gradient Boosting*), Adaboost e XGBoost e, cada um, conta com uma série específica de hiperparâmetros. Inicialmente, os hiperparâmetros padrão foram considerados e, para melhorar os modelos, foi realizada uma etapa de ajuste fino utilizando a função GridSearchCV com posterior avaliação por validação cruzada.

Os parâmetros *Boosting* utilizados para o ajuste final de cada modelo podem ser verificados na Tabela 10.

Tabela 10 - Hiperparâmetros *Boosting*, modelos de Carbamazepina

Caso	Algoritmo	Hiperparâmetros ajustados
1.000 ng. L ⁻¹	Gradient Boosting	'max_depth': 3, 'max_leaf_nodes': 4, 'n_estimators': 19
	Adaboost	'learning_rate': 0.1, 'n_estimators': 60
	XGboost	'learning_rate': 0.001, 'n_estimators': 50
500 ng. L ⁻¹	Gradient Boosting	'max_depth': 3, 'max_leaf_nodes': 8, 'n_estimators': 28
	Adaboost	'learning_rate': 0.1, 'n_estimators': 60
	XGboost	'learning_rate': 0.01, 'max_depth': 6, 'n_estimators': 30
100 ng. L ⁻¹	Gradient Boosting	'max_depth': 6, 'max_leaf_nodes': 9, 'n_estimators': 28
	Adaboost	'learning_rate': 0.001, 'n_estimators': 10
	XGboost	'learning_rate': 0.01, 'max_depth': 9, 'n_estimators': 10

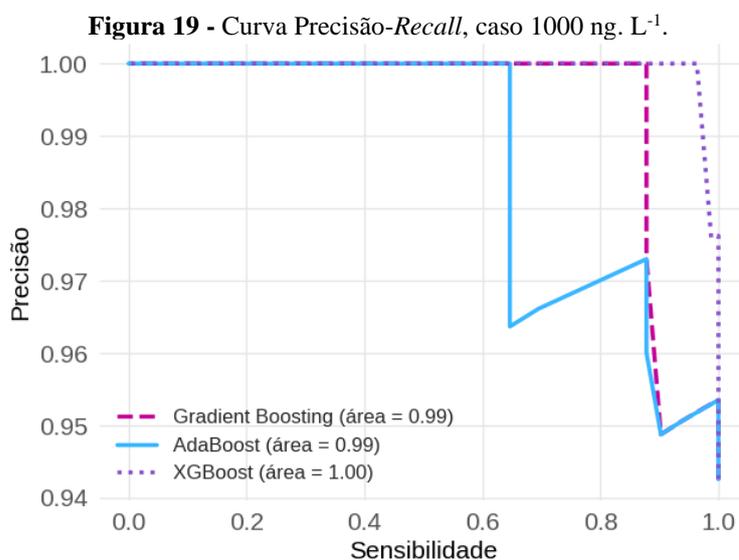
A comparação da Acurácia obtida pelos modelos de cada caso é apresentada graficamente na Figura A.1 do Apêndice A – Avaliação dos modelos. É importante evidenciar que para todos os casos, a construção dos modelos contou com a realização do balanceamento

de classes do conjunto de treinamento por meio do método SMOTE. O resultado do balanceamento para os limites de corte estabelecidos é exibido na Tabela 11.

Tabela 11 - Balanceamento de dados, modelos de Carbamazepina

Caso	Treinamento	Teste	Treinamento - SMOTE
1.000 ng. L ⁻¹	1: 317 0: 30	1: 82 0: 5	1: 317 0: 317
500 ng. L ⁻¹	1: 258 0: 89	1: 71 0: 16	1: 258 0: 258
100 ng. L ⁻¹	1: 61 0: 286	1: 16 0: 71	1: 286 0: 286

Para o caso de 1.000 ng. L⁻¹, a classe 1 (Aceitável) indica os valores de concentração de saída igual ou abaixo de 1.000 ng. L⁻¹, enquanto a classe 0 (Preocupante) compreende os valores acima de 1.000 ng. L⁻¹. A classe 1 é majoritária para os casos de 1.000 ng. L⁻¹ e 500 ng. L⁻¹ e, para garantir a representatividade de ambas as classes no ajuste, os dados de treinamento da classe 0 foram sobreamostrados para a mesma quantidade de observações da classe 1. Para o caso de 100 ng. L⁻¹, cuja classe majoritária é a classe 0, o balanceamento compreendeu a reamostragem dessa classe. Na Figura 19 verifica-se as curvas Precisão-Recall para os modelos ajustados.



Observa-se que os três modelos obtiveram alta precisão e alto *Recall* (Sensibilidade), com desempenho semelhante. A legenda da figura mostra que o XGBoost e Boosting Gradiente

demonstraram possuir a maior Área Sob a Curva (*Area Under the Curve*, AUC), o que indica que apresentam alto desempenho na classificação de amostras positivas (classe aceitável).

Na Tabela 12 é possível verificar as métricas de avaliação para os ajustes de 1.000 ng. L⁻¹, bem como o tempo de execução do treinamento.

Tabela 12 - Avaliação de modelos – 1.000 ng. L⁻¹

Algoritmo	Acurácia	F1 – score (Macro avg)	Tempo de treinamento (s)
Gradient Boosting	0,66	0,52	30,69
Adaboost	0,64	0,47	4,63
XGboost	0,80	0,63	0,30

O Adaboost e o XGBoost apresentaram tempo de treinamento menor do que o Gradient Boosting, mesmo com maior quantidade de estimadores (60 e 50 estimadores, respectivamente). O Adaboost constrói árvores pequenas, com apenas 1 nó raiz e 2 folhas e, o XGBoost, embora seja flexível à construção de árvores de maior profundidade, possui uma série de aperfeiçoamentos que garantiram sua maior velocidade.

O algoritmo XGBoost apresentou métricas com melhor desempenho em comparação ao Adaboost e Gradient Boosting. O valor de acurácia obtido por esse algoritmo foi de 0,80, enquanto os modelos Adaboost e Gradient Boosting atingiram valores de 0,64 e 0,66. Portanto, o XGBoost apresentou maior taxa de acerto nas classificações considerando ambas as classes.

O mesmo foi verificado para os valores de *F1 - score*. Essa métrica combina a precisão e *Recall*, sendo que os valores mais próximos de 1 indicam melhor desempenho. Neste caso, o *F1 - score Macro Averaged* agrega os valores de *F1 - score* de cada classe, sem ponderar os resultados caso haja desbalanceamento, de forma a garantir que ambas as classes apresentem a mesma importância. Os valores obtidos de *F1 - score* em contraste com a Acurácia indicam que o Recall ou a Precisão das classes, individualmente, apresentou valores reduzidos, o que levou a valores de F1-score menores do que a acurácia.

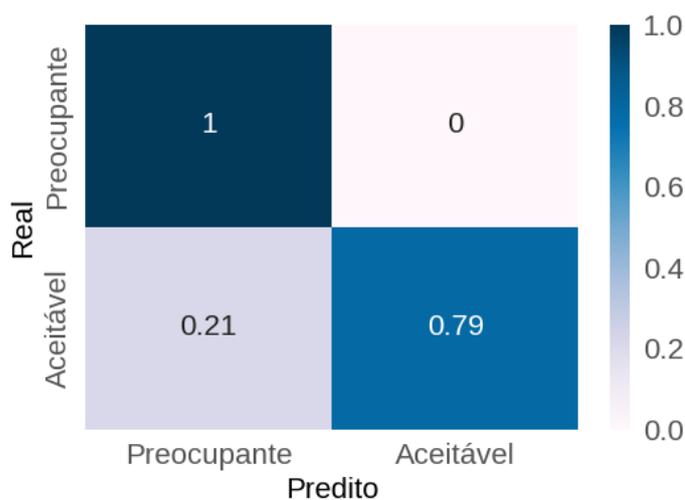
O balanceamento das classes se deu apenas para o conjunto de treinamento e, de 87 observações do conjunto de teste, 82 pertenciam a classe 1 (aceitável) e 5 pertenciam a classe 0 (preocupante). Com isso, os modelos apresentaram baixa precisão na classificação de observações de rótulo “preocupante”. As métricas obtidas para cada classe a partir do modelo XGBoost são indicadas na Tabela 13.

Tabela 13 - Métricas XGBoost – 1.000 ng. L⁻¹

Classe	Precisão	Recall	F1 - score	Observações
0 (preocupante)	0,23	1	0,37	5
1 (aceitável)	1	0,79	0,88	82

Os valores de F1 – score obtidos para cada classe revelaram uma diferença considerável e confirmam que o modelo foi capaz de prever de forma mais assertiva as concentrações da classe “aceitável” (F1 = 0,88) em relação à classe “preocupante” (F1 = 0,37). Dentre as observações classificadas como “aceitável” o modelo foi capaz de acertar 100% (Precisão = 1). Considerando todas as observações que de fato são da classe “aceitável” o modelo classificou corretamente 79%. Na Figura 20 é possível verificar a matriz de confusão deste modelo.

Figura 20 - Matriz de Confusão normalizada do modelo XGboost, caso 1.000 ng. L⁻¹

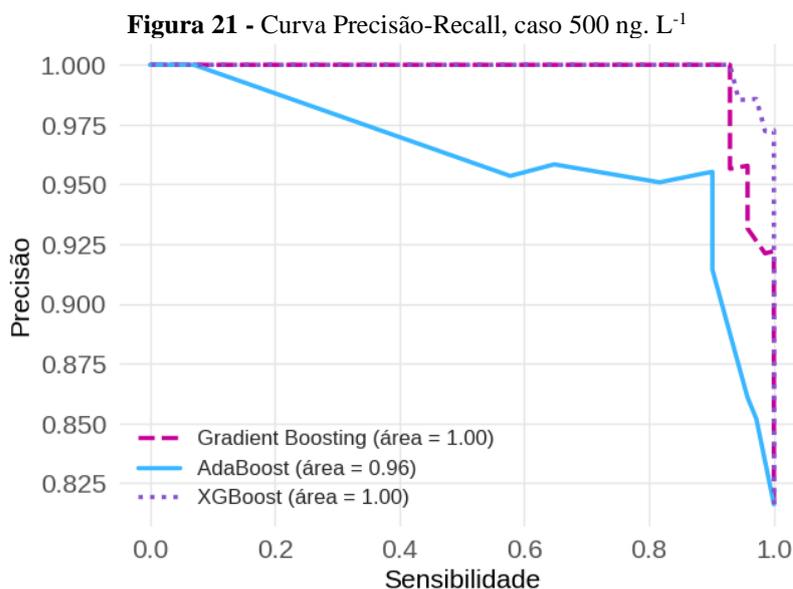


Fonte: Elaborado pela autora.

Apenas 23% de precisão foi obtida para a classe “preocupante”, visto que 21% de instâncias da classe “aceitável” foram erroneamente classificadas nessa classe. Apesar disso, todas as observações reais da classe “preocupante” foram classificadas corretamente pelo modelo, o que justifica o *Recall* de valor 1.

O maior valor de *F1 - score* geral (0,63) foi obtido pelo XGBoost e corrobora para que este algoritmo seja considerado o de maior capacidade preditiva, quando comparado aos outros dois algoritmos testados. Este algoritmo gerou o modelo com menor taxa de falsos negativos (17 instâncias) e maior taxa de verdadeiros positivos (65 instâncias) e foi o que apresentou maior precisão na classificação de amostras negativas.

Para os casos de limite de rotulação de 500 ng. L⁻¹, foi garantida uma maior distribuição das observações entre as classes. No entanto, como o balanceamento foi aplicado apenas ao conjunto de treinamento, foi possível verificar a predominância de observações na classe 1 no conjunto de teste, como indicado na Tabela 11. A curva de Precisão-*Recall* para os modelos é apresentada na Figura 21.



Os modelos XGBoost e Gradient Boosting apresentaram a maior área sob a curva (AUC = 1,0) e, portanto, demonstraram o melhor desempenho. O modelo Adaboost também apresentou alto valor de área sob a curva, mas foi superado pelos outros ajustes.

Na Tabela 14 são apresentadas as métricas de avaliação e o tempo de execução do treinamento para cada algoritmo. Novamente, mesmo que os algoritmos Adaboost e XGBoost tenham sido executados com maior quantidade de estimadores do que o Gradient Boosting, ambos apresentaram um treinamento rápido.

Tabela 14 - Avaliação de modelos – 500 ng. L⁻¹

Algoritmo	Acurácia	F1 – score (macro)	Tempo de treinamento (s)
Gradient Boosting	0,90	0,85	44,92
Adaboost	0,88	0,82	3,91
XGboost	0,94	0,91	6,01

Embora o Adaboost tenha demonstrado menor tempo de execução do treinamento, foi o modelo com menor acurácia, sendo superado pelo Gradient Boosting e XGBoost, que apresentaram acurácia de 0,90 e 0,94, respectivamente. Os valores de *F1 - score Macro Averaged* foram 0,85, 0,82 e 0,91 para o Gradient Boosting, Adaboost e XGBoost, respectivamente. Isso implica que obtiveram desempenho satisfatório não apenas para a classe majoritária (Aceitável), mas também, para a classe minoritária (Preocupante).

Assim, o nível de desbalanceamento verificado no conjunto de teste teve menor impacto negativo na precisão e *Recall* para o limite intermediário de 500 ng. L⁻¹ em comparação ao caso

apresentado anteriormente. Para o caso atual, o XGboost apresentou maior acurácia e *F1 - score*. O desempenho obtido para cada classe pode ser avaliado na Tabela 15.

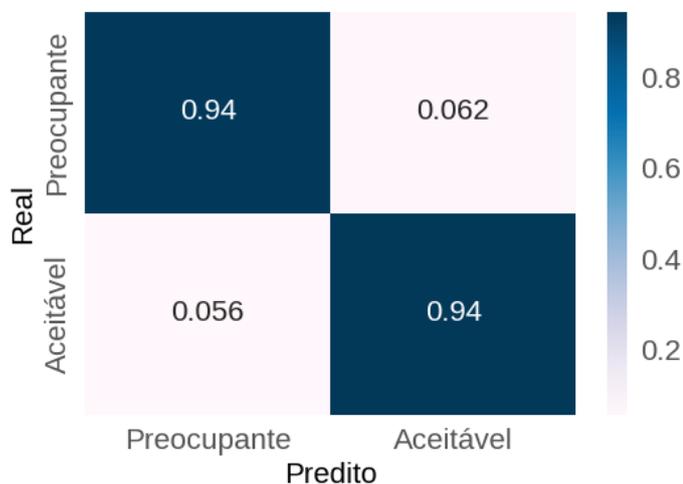
Tabela 15 - Métricas XGBoost – 500 ng. L⁻¹

	Precisão	Recall	F1 - score	Observações
Classe 0	0,79	0,94	0,86	16
Classe 1	0,99	0,94	0,96	71

Verifica-se que a classe majoritária apresenta precisão de 99%, o que indica que o modelo apresentou uma maior taxa de acerto para as instâncias classificadas como “aceitável”, enquanto para a classe minoritária o índice de acerto foi de 79%. O *Recall* indica que o modelo foi capaz de classificar corretamente 94% das observações reais pertencentes à classe “preocupante” e 94% das observações reais pertencentes à classe “aceitável”.

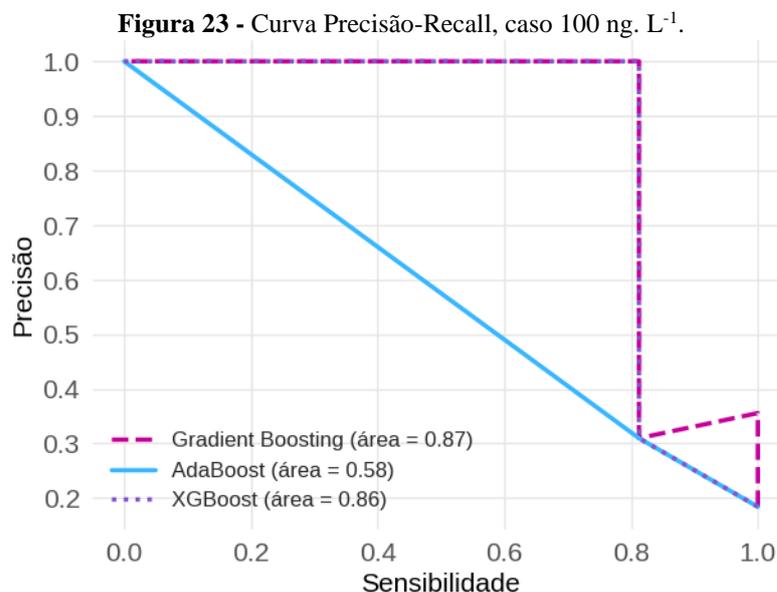
Pela matriz de confusão exibida na Figura 22, é possível verificar que 6,2% das instâncias da classe preocupante (0) foram classificadas como aceitáveis (1) e 5,6% de instâncias da classe aceitável foram classificadas como preocupante, o que justifica a menor precisão da classe “preocupante” em comparação à classe “aceitável”.

Figura 22 - Matriz de Confusão normalizada do modelo XGBoost, caso 500 ng. L⁻¹.



Fonte: Elaborado pela autora.

Para o caso de limite de corte de 100 ng. L⁻¹, a mesma situação dos outros limites foi verificada para o balanceamento das classes do conjunto de teste. Isso impactou o resultado das métricas de avaliação calculadas a partir deste conjunto. Na Figura 233 verifica-se a curva *Precisão-Recall*.



O modelo Adaboost apresentou menor área sob a curva, o que reflete seu menor desempenho. Já os modelos Gradient Boosting e XGBoost apresentaram valores de AUC equivalentes a 0,87 e 0,86, demonstrando que tiveram maior capacidade de acertos. Na Tabela 16 é possível verificar as métricas de avaliação, bem como o tempo de execução do treinamento de cada modelo.

Tabela 16 - Avaliação de modelos – 100 ng. L⁻¹

Algoritmo	Acurácia	F1 – score (macro average)	Tempo de treinamento (s)
Gradient Boosting	0,89	0,84	64,12
Adaboost	0,63	0,59	3,49
XGboost	0,94	0,90	4,12

O XGboost apresentou maior acurácia e *F1 – score* e o Adaboost apresentou desempenho inferior aos outros, que pode ser justificado pela menor quantidade de estimadores (10) empregada. Já o Gradient Boosting apresentou desempenho intermediário e, tendo ajustado árvores de maior profundidade (profundidade máxima = 9) e maior quantidade de estimadores (28), enquanto o XGBoost obteve o maior desempenho utilizando árvores de profundidade máxima 9 e utilizando 10 estimadores.

Para esse caso, a classe 0 (preocupante) foi a majoritária, contendo 71 observações. Já a classe 1 (aceitável) apresentou 16 observações. Isso ocorreu porque o limite de 100 ng. L⁻¹ é mais restritivo do que os anteriores e, portanto, menos observações atenderam ao critério de

rotulação que estabeleceu como concentração de saída aceitável apenas observações menores do que 100 ng. L⁻¹. Na Tabela 17 é possível verificar as métricas para ambas as classes obtidas pelo modelo XGBoost.

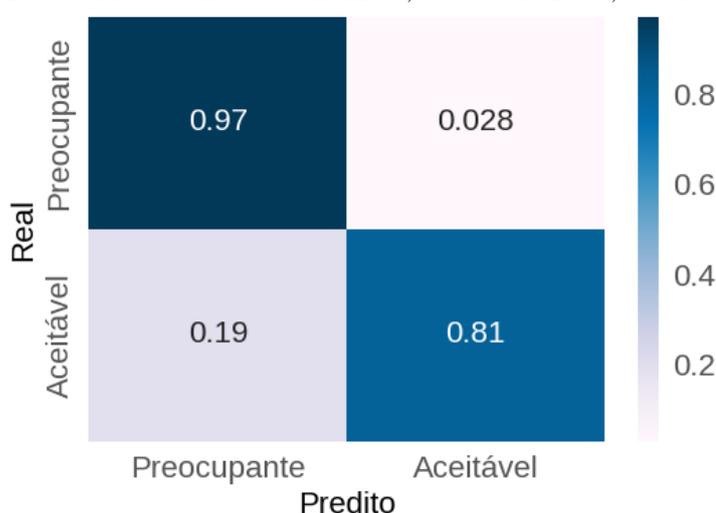
Tabela 17 - Métricas XGBoost – 100 ng. L⁻¹

	Precisão	Recall	F1 - score	Observações
Classe 0	0,96	0,97	0,97	71
Classe 1	0,87	0,81	0,84	16

A classe majoritária obteve maior precisão, *Recall* e *F1 - score*, confirmando que o modelo apresentou melhor desempenho para classificar observações da classe “preocupante”. Ao passo que para a classe “aceitável”, minoritária, o modelo obteve menor desempenho. Ainda assim, apresentou 87% de acerto na classificação e entre todas as observações que realmente integram a classe 1, o modelo foi capaz de classificar corretamente 81%, atingindo um *F1 - score* de 84%.

Desse modo, o modelo XGBoost foi capaz de prever satisfatoriamente ambas as classes. Pela matriz de confusão exibida na Figura 244, é possível verificar que 81% das amostras foram classificadas corretamente como positivas (classe aceitável, 1), enquanto houve uma taxa de 19% de falsos negativos. Já a classe negativa (preocupante, 0) compreendeu 97% de taxa de verdadeiros negativos, amostras negativas corretamente classificadas, e 2,8% de falsos positivos, amostras erroneamente classificadas como positivas.

Figura 24 - Matriz de Confusão normalizada, Modelo XGBoost, caso 100 ng. L⁻¹.



Fonte: Elaborado pela autora.

Os resultados demonstram que para a abordagem preditiva do nível de concentração final de Carbamazepina, os algoritmos XGBoost e Gradient Boosting apresentaram melhor

desempenho para os limites de corte investigados, sendo que o XGBoost foi o de maior desempenho. Em nenhum dos casos apresentados o algoritmo Adaboost foi capaz de superar os outros.

Esse aspecto pode indicar que embora seja um método eficiente, sua estrutura baseada no treinamento de árvores pequenas foi meramente satisfatória, e quando comparada aos outros algoritmos utilizados sua capacidade preditiva demonstrou ser insuficiente, não apresentando maior destaque para os casos cujos dados de treinamento apresentam alta variabilidade.

O limite de corte permissivo de 1.000 ng. L⁻¹ apresentou a maior discrepância entre os valores de precisão e, conseqüentemente, *F1 - score* das classes. Foi observado o maior desbalanceamento entre as classes no conjunto de teste, com apenas 5 observações da classe “preocupante”.

Nesse caso, o algoritmo XGBoost foi o que apresentou a maior capacidade de classificação de ambas as classes. No entanto, ao se avaliar o *F1 - score* para cada classe, o valor de 0,37 obtido para a classe minoritária (classe 0: preocupante) demonstrou que mesmo um algoritmo potente não foi satisfatório para prever adequadamente as situações de concentração preocupante considerando um limite de rotulação tão extremo.

Para os casos de limites de corte de 500 ng. L⁻¹ e 100 ng. L⁻¹, os modelos foram capazes de classificar corretamente e apresentar alta taxa de captura dos exemplos reais de ambas as classes. O caso de limite de corte de 100 ng. L⁻¹ foi o mais restritivo e, levando em conta a faixa de Concentração de saída (ng. L⁻¹) verificada para a Carbamazepina, justifica-se o fato de que a classe majoritária tenha passado a ser a classe “Preocupante” (0).

A comparação entre os modelos de maior desempenho em cada nível de corte é mostrada na Tabela 18.

Tabela 18 - Comparação entre modelos em diferentes limites de classificação

Nível de corte	Técnica	F1 - score macro averaged	Considerações
1.000	XGBoost	0,63	Poucos exemplos da classe “Preocupante” (0). O modelo não foi capaz de classificar satisfatoriamente amostras dessa classe.
500	XGBoost	0,91	O modelo foi capaz de classificar satisfatoriamente ambas as classes, mesmo com uma menor quantidade de amostras da classe minoritária (Preocupante, 0).
100	XGBoost	0,90	O modelo foi capaz de classificar satisfatoriamente ambas as classes, mesmo com uma menor quantidade de amostras da classe minoritária (Aceitável, 1).

Embora para todos os casos tenha sido observada uma discrepância entre a quantidade de exemplos das classes nos dados de teste, o limite de corte de 1.000 ng. L⁻¹ se apresentou como o mais desafiador em termos de gerar um modelo admissível. Nesse caso, trabalhar com um limite de rotulação mais permissivo não se mostrou compatível ao perfil das concentrações modeladas. Por isso, para a abordagem de Carbamazepina esse caso não apresentou uma contribuição expressiva.

Em contrapartida, os modelos construídos para os casos de 500 ng. L⁻¹ e 100 ng. L⁻¹, demonstraram uma melhor capacidade preditiva, já que mesmo diante de discrepâncias entre a quantidade de observações de cada classe no conjunto de teste, em ambos os casos se verificou um alto desempenho. É possível avaliar que tanto o limite de corte intermediário como o limite restritivo proporcionaram resultados mais robustos para o conjunto de dados utilizados.

A Carbamazepina é um princípio ativo altamente persistente, que, normalmente, atinge baixos valores de remoção (aproximadamente 30%), como verificado previamente (KOT-WASIK; JAKIMSKA; ŚLIWKA-KASZYŃSKA, 2016; SUBEDI et al., 2017; VERGILI et al., 2019). Esse composto também se destaca pela alta tendência em formar metabólitos tóxicos, o que evidencia um perfil complexo de degradação, que possivelmente, é um dos fatores que pode acarretar em incertezas nas predições, como notado por Verlicchi et al. (2014).

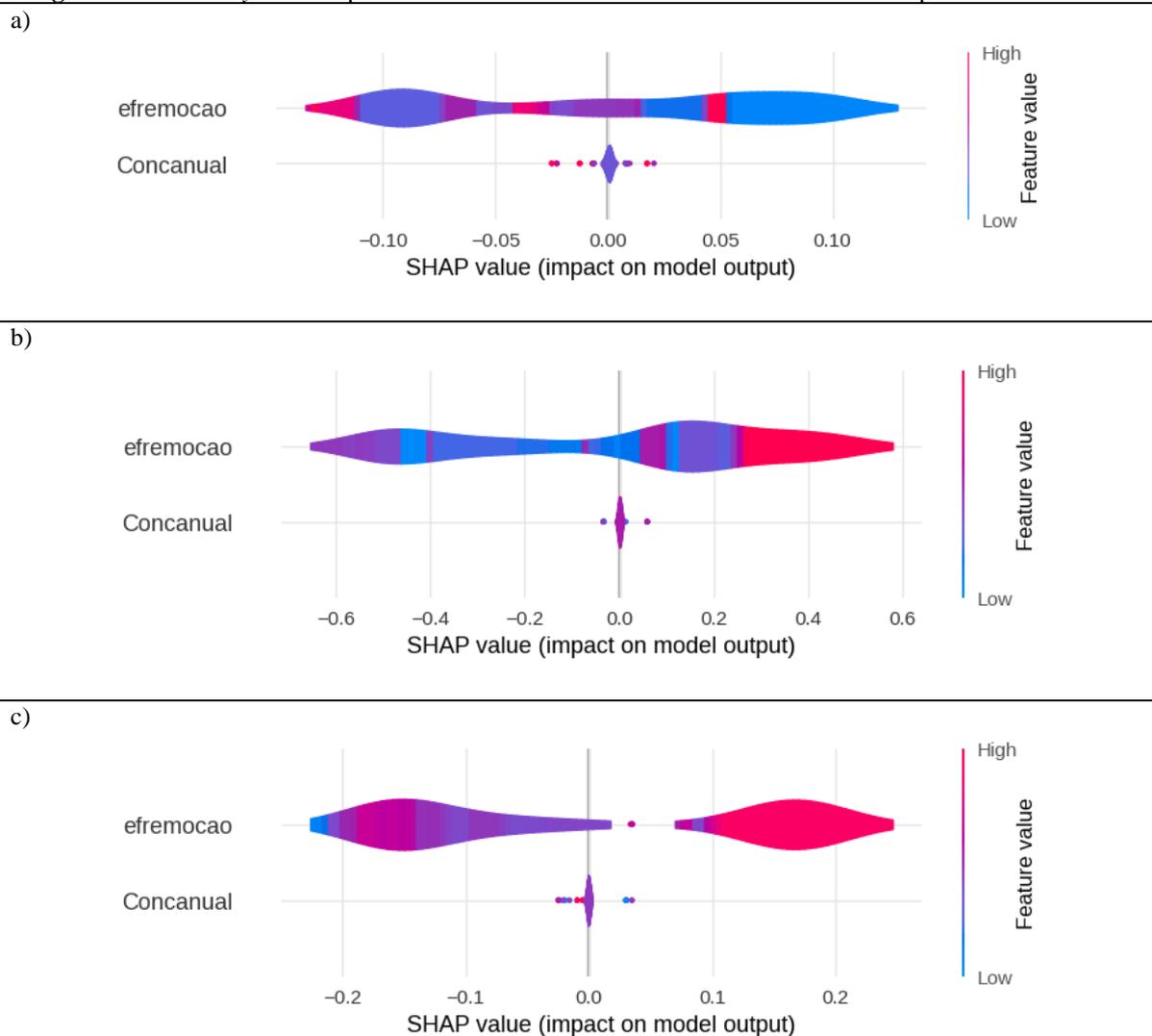
Casas et al. (2024) utilizaram o modelo Random Forest para a classificação de taxas de remoção de fármacos em efluentes com diferentes tipos de solventes. A acurácia dos modelos se manteve acima de 91%, exceto para a Carbamazepina, que apresentou o valor 87% para o conjunto de teste. Esse princípio ativo foi identificado como o de menor previsibilidade. Os autores atribuíram esse fato possivelmente à inclusão do preditor pH, que não apresentou influência relevante no processo.

Mesmo diante do perfil comportamento destoante da Carbamazepina, os modelos *boosting* aplicados para os limites de rotulação de 500 ng. L⁻¹ e 100 ng. L⁻¹ demonstraram acurácia superior ao resultado encontrado por Casas et al. (2024). Já para o limite de 1000 ng. L⁻¹, os modelos apresentaram desempenho inferior, o que pode sugerir que outras variáveis poderiam ser relevantes para a classificação.

5.4.1.2 Importância das variáveis

Na Figura 25 é possível verificar o resumo da contribuição das variáveis na classificação das observações. Mostra-se, para cada caso, os resultados do modelo de melhor desempenho.

Figura 25 - Summary Plot: Impacto das variáveis na saída do modelo de maior desempenho de cada nível



a) XGBoost – 1.000 ng. L⁻¹, b) XGBoost – 500 ng. L⁻¹, c) XGBoost – 100 ng. L⁻¹. *efremocao: Eficiência de Remoção de Carbamazepina; Concanual: Concentração anual de Carbamazepina comercializada.

No eixo y, as variáveis são dispostas conforme o nível de importância, com a de maior importância ocupando a parte superior seguida pelas variáveis de menor importância. Para cada variável existe um valor shap associado à resposta do modelo (classe de concentração) para cada instância e, portanto, como o shap foi implementado para os dados de treinamento, para cada variável plotada há 434 valores shap distribuídos no eixo x.

Assim, no eixo x é possível verificar a magnitude do impacto das variáveis (valor shap), bem como a distribuição desses valores. Quanto maior a expansão nesse eixo, maior é o impacto observado para aquela variável no modelo e seu poder preditivo pode ser constatado. Além disso, a cor indica o valor da variável, sendo a tonalidade azulada corresponde a valores baixos e a avermelhada relacionada a valores altos. Nesse sentido, a escala de cores possibilita

compreender de que forma os valores baixos ou altos das variáveis impactam na classe final das observações.

Para todos os casos, verifica-se que a eficiência de remoção (Eficiência de remoção) foi a variável mais importante para o modelo e a que apresentou maior nível de impacto nas saídas individuais das observações. A concentração anual de Carbamazepina (Concanual) que atinge a ETE para tratamento, notadamente teve menor importância e impacto baixíssimo.

A Figura 25(a) indica que valores baixos de eficiência de remoção se associam à resposta da classe positiva (aceitável). No entanto, o SHAP negativo encontra-se associado à escala de tons de roxo da variável Eficiência de remoção indicando que valores considerados medianos aumentam a chance de se obter como resposta a classe de concentração de saída “preocupante”.

Esse comportamento contraintuitivo pode estar associado ao fato de que mesmo em casos em que se pode ter atingido alta eficiência de remoção para Carbamazepina, a concentração de saída no tratamento ainda foi igual ou superior ao limite extremo para rotulação das classes: 1.000 ng. L^{-1} . A Carbamazepina demonstra ser um princípio ativo altamente resistente aos tratamentos de remoção, e nos dados utilizados para a modelagem apresentou eficiência de remoção que variou de 0,28% a 94,90%, com média de 18,27%.

Portanto, o *summary plot* indicou que o modelo ajustado para um alto limite de rotulação pode ter identificado casos extremos de dados de alta concentração de ocorrência, em que a elevada eficiência de remoção obtida ainda não foi capaz de reduzir a concentração de saída para um valor abaixo do limite de rotulação das classes (1.000 ng. L^{-1}). Dessa forma, o modelo pode ter interpretado que esses casos aumentam o impacto para se obter a classe de saída preocupante.

Ao mesmo tempo, o modelo identificou casos com valores mais baixos de concentração de ocorrência, em que mesmo com eficiência de remoção reduzidas, o valor de concentração de saída se manteve abaixo do limite de corte definido. Possivelmente, esses casos estão associados ao impacto para ocorrência da classe aceitável.

Na Figura 25(b) é possível verificar uma ampla distribuição de valores reduzidos de eficiência de remoção que impactam, em sua maioria, para a obtenção de uma resposta final da classe negativa, de concentração preocupante. É possível distinguir uma região onde verifica-se eficiência de remoção elevada gerando impacto positivo na saída do modelo, aumentando as chances de se obter a classe de concentração aceitável.

A Figura 25(c) exibe que para o modelo ajustado no limite de corte de 100 ng. L^{-1} é possível identificar que valores reduzidos e medianos de eficiência de remoção estão associados

a SHAP negativos enquanto valores elevados estão associados a SHAP positivos, corroborando que valores superiores de eficiência de remoção aumentam as chances de se obter uma concentração de saída aceitável.

Desse modo, foi possível identificar o impacto das variáveis na classificação considerando diferentes níveis para o estabelecimento das classes. Foi constatado que o modelo foi impactado em maior nível pela Eficiência de remoção, enquanto a concentração de entrada em ETE proveniente dos dados de comercialização não exerceu impacto nas decisões do modelo. É importante destacar que os dados de Carbamazepina possuem valores extremos de concentrações de ocorrência, além de alta dispersão de eficiência de remoção, onde predominam valores reduzidos. Isso porque a Carbamazepina é um insumo farmacêutico ativo altamente resistente aos tratamentos de remoção empregados, o que gera um perfil com bastante variabilidade nos dados.

5.4.2 Modelos para os Antibióticos

5.4.2.1 Métricas de Avaliação

Construíram-se ajustes para classificar a concentração final de antibióticos em três níveis para rotulação das classes “aceitável” e “preocupante”. Os algoritmos Gradient Boosting Model, Adaboost e XGBoost foram implementados e os hiperparâmetros foram obtidos pelo método *GridSearchCV*. Este método faz uma busca exaustiva sobre valores de hiperparâmetros para um dado estimador de forma otimizada por um processo de validação cruzada. Os hiperparâmetros utilizados em cada caso estão explicitados na Tabela 19.

Tabela 19 - Hiperparâmetros Boosting para modelos de Antibióticos

Caso	Algoritmo	Hiperparâmetros ajustados
1.000 ng. L ⁻¹	Gradient Boosting	'learning_rate': 0.01, 'max_depth': 2, 'max_leaf_nodes': 2, 'min_samples_leaf': 1, 'n_estimators': 1
	Adaboost	'learning_rate': 1e-05, 'n_estimators': 1
	XGboost	'learning_rate': 0.1, 'max_depth': 1, 'n_estimators': 60
500 ng. L ⁻¹	Gradient Boosting	'max_depth': 6, 'max_leaf_nodes': 8, 'n_estimators': 141
	Adaboost	'learning_rate': 1, 'n_estimators': 65
	XGboost	'learning_rate': 1, 'max_depth': 3, 'n_estimators': 100
100 ng. L ⁻¹	Gradient Boosting	'learning_rate': 0.01, 'max_depth': 8, 'max_leaf_nodes': 9, 'n_estimators': 40
	Adaboost	'learning_rate': 0.1, 'n_estimators': 45

XGboost

'learning_rate': 0.01, 'max_depth': 8, 'n_estimators': 40

A técnica SMOTE foi utilizada para o balanceamento dos dados para os ajustes nos limites de 1.000 ng. L⁻¹ e 500 ng. L⁻¹. Para o caso de 100 ng. L⁻¹ não foi necessário realizar o balanceamento, já que as observações estavam bem distribuídas entre as classes, como pode ser observado na Tabela 20.

Tabela 20 - Balanceamento de classes para modelos de Antibióticos

Caso	Treinamento	Teste	SMOTE (treinamento)	SMOTE (teste)
1.000 ng. L ⁻¹	1: 381 0: 6	1: 165 0: 1	1: 381 0: 381	1: 165 0: 165
500 ng. L ⁻¹	1: 322 0: 65	1: 150 0: 26	1: 322 0: 322	-
100 ng. L ⁻¹	1: 207 0: 180	1: 94 0: 72	-	-

No caso de 1.000 ng. L⁻¹, o balanceamento foi realizado tanto nos dados de treinamento quanto nos de teste, visto que originalmente o conjunto de teste continha apenas 1 observação da classe preocupante. Assim, com os dados balanceados, os modelos foram ajustados para esse caso.

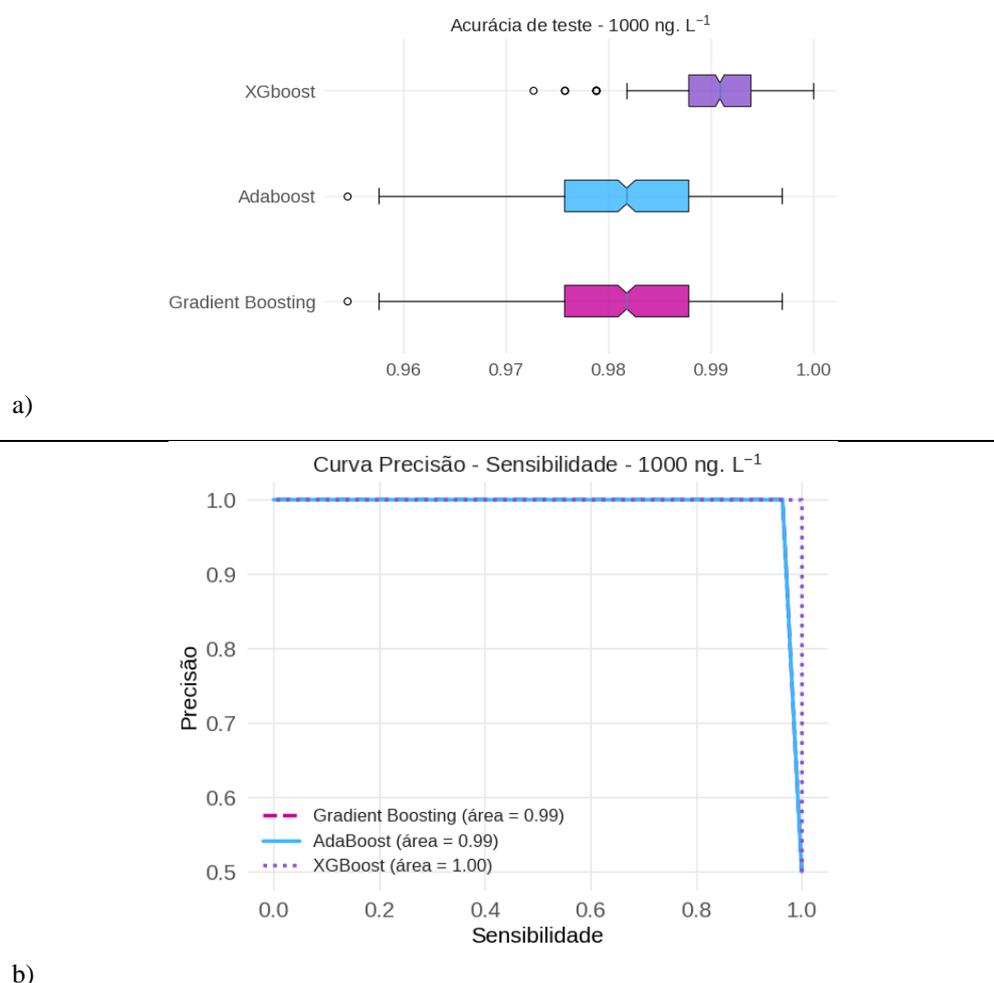
Esse limite é o mais permissivo, pois considera “preocupante” apenas as observações com concentração de saída igual ou acima do valor 1.000 ng. L⁻¹. Na Tabela 21 verifica-se o desempenho obtido nos dados de teste em cada ajuste.

Tabela 21 - Avaliação de modelos – 1000 ng. L⁻¹

Algoritmo	Acurácia	F1 – score (Macro avg)	Tempo de treinamento (s)
Gradient Boosting	0,98	0,98	2,35
Adaboost	0,98	0,98	0,32
XGboost	0,99	0,99	14,86

Os modelos apresentaram acurácia e *F1 - score* similares, sendo que o XGBoost apresentou valores superiores. No entanto, sua execução demorou 14,29 segundos, enquanto o Adaboost e Gradient Boosting apresentaram menor tempo de execução. Isso pode ser consequência do fato de que foram utilizados 60 estimadores para o XGBoost.

A comparação entre o desempenho dos modelos pode ser verificada na Figura 266.

Figura 26 - Desempenho de modelos, caso 1.000 ng. L⁻¹

b) Acúrcia de teste; b) Curva Precisão-Recall, para modelos de 1.000 ng. L⁻¹.
 Fonte: Elaborado pela autora.

Na Figura 26(a) é possível notar que embora a acúrcia entre os modelos seja próxima, o XGBoost se destaca com um valor em torno de 0,99. Seu desempenho superior também é verificado na Figura 26(b) que demonstra o valor de AUC 1,0. Desse modo, embora todos os ajustes tenham apresentado alto desempenho o XGBoost demonstrou ser superior nos acertos da classificação, atingindo *F1 - score* de 0,99 para ambas as classes. A matriz de confusão desse ajuste pode ser verificada na Figura A.2a, no Apêndice A – Avaliação dos modelos.

Para o caso de 500 ng. L⁻¹, os modelos também apresentaram alto desempenho. O gráfico comparativo da Acúrcia pode ser visualizado na Figura A.3a, no Apêndice A e, as métricas podem ser verificadas na Tabela 22.

Tabela 22 - Avaliação de modelos – 500 ng. L⁻¹

Algoritmo	Acúrcia	F1 – score (Macro avg)	Tempo de treinamento (s)
-----------	---------	------------------------	--------------------------

Gradient Boosting	0,95	0,91	2863,07
Adaboost	0,93	0,88	26,51
XGboost	0,94	0,90	68,42

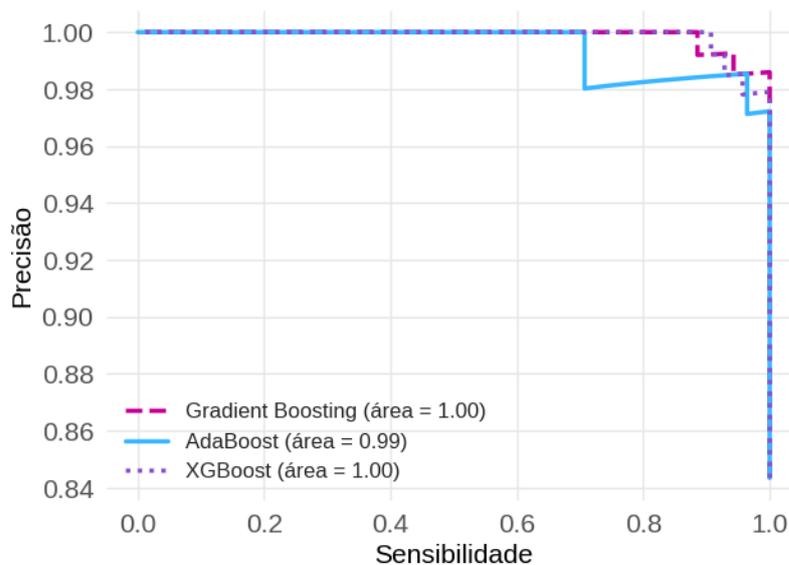
O modelo Gradient Boosting apresentou maior acurácia e maior *F1 - score*. No entanto, sua execução levou aproximadamente 47 minutos, um tempo longo em comparação aos outros modelos. Isso pode ter relação com o alto tempo de busca dos hiperparâmetros ótimos, que implicou como resultado final na geração de um modelo com maior profundidade das árvores (profundidade máxima 6), maior tamanho das folhas (tamanho máximo de amostras 8) e mais quantidade de estimadores (141), levando a um maior tempo de treinamento.

Desse modo, o ajuste realizado com o XGBoost demonstrou o melhor custo-benefício devido ao tempo de execução intermediário, e também, os altos valores de acurácia e *F1 - score*. O ajuste baseado no modelo Adaboost também apresentou desempenho satisfatório, mas menor do que os outros ajustes. Isso pode ter ocorrido devido ao fato desse modelo ter utilizado apenas 65 estimadores, enquanto os outros modelos utilizaram 141 (Gradient Boosting) e 100 (XGBoost) estimadores. Além disso, o Adaboost é composto por *stumps* de apenas 1 nível de profundidade. Por isso, apresentou o menor desempenho entre os ajustes e, sua execução normalmente é mais rápida.

A curva de Precisão-*Recall* exibida na

Figura 277 demonstra que todos os ajustes apresentaram alta taxa de recuperação de amostras positivas. No entanto, o perfil do Gradient Boosting e XGBoost se destacam pela semelhança e mesmo valor de AUC.

Figura 27 - Curva Precisão-Recall, caso 500 ng. L⁻¹



Fonte: Elaborado pela autora.

Na Tabela 23 mostra-se a precisão, *Recall*, *F1 - score* e quantidade de observações de teste para cada classe com base no modelo XGBoost, de melhor custo-benefício.

Tabela 23 - Métricas XGBoost 500 ng. L⁻¹

Classe	Precisão	Recall	F1 - score	Observações
0 (preocupante)	0,79	0,88	0,83	26
1 (aceitável)	0,97	0,95	0,96	140

O modelo foi capaz de distinguir bem ambas as classes, mesmo a classe 0 (preocupante) que apresentou a menor quantidade de dados de teste. Embora as métricas da classe “preocupante” tenham sido inferiores as métricas da classe 1, os valores foram satisfatórios e mostram que o XGBoost apresentou alta capacidade de decisão. O modelo foi capaz de acertar 79,3% das classificações e, conseguiu capturar 88,5% das observações que realmente pertenciam a essa classe. A matriz de confusão do XGBoost pode ser verificada na Figura A.2b, no Apêndice A – Avaliação dos modelos. O resultado obtido pode ser considerado vantajoso dado que essa classe é minoritária no conjunto de teste.

Para os ajustes realizados para o limite de corte de 100 ng. L⁻¹, a Tabela 24 exhibe os valores das métricas de desempenho e o tempo de treinamento. É importante evidenciar que nesse caso, as observações estavam consideravelmente balanceadas entre as classes e, por isso, não foi realizado qualquer procedimento de balanceamento de classes.

Tabela 24 - Avaliação de modelos – 100 ng. L⁻¹

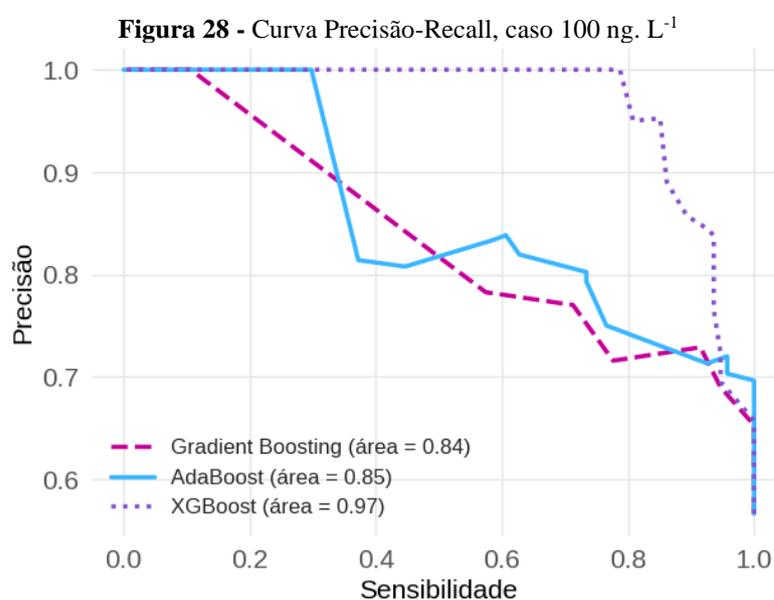
Algoritmo	Acurácia	F1 - score (Macro avg)	Tempo de treinamento (s)
Gradient Boosting	0,69	0,68	165,37

Adaboost	0,72	0,71	5,71
XGboost	0,86	0,85	8,19

O algoritmo Gradient Boosting apresentou o maior tempo de execução, considerando a busca por hiperparâmetros e o treinamento das árvores com profundidade máxima igual a 8, quantidade máxima de observações em uma folha igual a 9 e 40 estimadores. O Adaboost apresentou desempenho intermediário e o menor tempo de execução entre os ajustes e, neste caso, foi necessário utilizar 45 estimadores para alcançar o desempenho apresentado.

O XGBoost apresentou a maior acurácia e *F1 - score*, com o tempo de treinamento intermediário em comparação aos outros algoritmos. Seu treinamento contou com a busca de hiperparâmetros e, ao final, o ajuste contou com profundidade máxima das árvores igual a 8 e 40 estimadores, assim como o Gradient Boosting. O XGBoost é uma otimização do Gradient Boosting e por isso, com as mesmas condições seu desempenho foi mais alto e sua execução mais rápida.

A comparação entre a Acurácia obtida por cada um dos algoritmos pode ser verificada na Figura A.3b, no Apêndice A, que mostra que o XGBoost atingiu maiores valores de acurácia de teste. A Figura 28 mostra a curva de Precisão-*Recall* para os ajustes desse limite de corte.



Fonte: Elaborado pela autora.

É possível notar que o Gradient Boosting e Adaboost obtiveram um perfil semelhante de detecção de amostras da classe aceitável. Já o algoritmo XGBoost se destacou por apresentar a maior área sob a curva, o que implica em geração de maior taxa de recuperação das amostras positivas. Seu desempenho por classe pode ser verificado na Tabela 25.

Tabela 25 - Métricas XGBoost 100 ng. L⁻¹

Classe	Precisão	Recall	F1 - score	Observações
0 (preocupante)	0,88	0,77	0,83	72
1 (aceitável)	0,84	0,92	0,88	94

O XGboost apresentou elevado desempenho para ambas as classes. O valor de *Recall* obtido para a classe 0 implica que 77,8% das instâncias de fato pertencentes a essa classe foi corretamente classificada pelo modelo, enquanto acertou 88,9% das classificações dessa classe. Já o *Recall* da classe 1 indica que 92,6% das instâncias pertencentes à classe aceitável foram corretamente detectadas pelo modelo. No entanto, dentre as classificações da classe 1, 84,5% foram corretas. A matriz de confusão pode ser visualizada na Figura A.2c, do Apêndice A.

Os resultados apontam que para o limite restritivo de 100 ng. L⁻¹ sem balanceamento de classes para o treinamento do ajuste, o XGBoost foi o método que notavelmente apresentou maior poder de classificação, sendo capaz de detectar e classificar adequadamente ambas as classes. Para todos os casos implementados, esse algoritmo demonstrou ser o de melhor desempenho e custo-benefício.

Para o caso da avaliação com os dados de concentração de antibióticos, não foi observada diferenças consideráveis entre o desempenho obtido em cada caso avaliado. Entretanto, para o limite de 1.000 ng. L⁻¹ foi realizado balanceamento das classes tanto para o conjunto de treinamento como o de teste, o que gerou uma condição de 50% de observações para cada classe em ambos os conjuntos.

Essa abordagem é válida para auxiliar a comparação entre diferentes formas de preparar os dados para a modelagem e avaliar o desempenho dos algoritmos. Porém, é importante evidenciar que casos reais raramente apresentam essa condição ideal de balanceamento. Por isso, os modelos desenvolvidos para o caso permissivo de 1.000 ng. L⁻¹ revelaram-se pouco factíveis e aplicáveis na prática.

Já o limite de corte de 500 ng. L⁻¹, inicialmente apresentou um desbalanceamento das observações de treinamento de: classe 1: 83%, classe 0: 17%. Assim, foi realizado o balanceamento apenas no conjunto de treinamento. Os modelos obtidos apresentaram alto desempenho e, o modelo XGBoost foi o de melhor custo-benefício, utilizando a maior quantidade de estimadores. Para esse caso, os modelos provaram-se mais adequados à realidade.

O limite de corte de 100 ng. L⁻¹ garantiu a partição de dados mais favorável para ambos os conjuntos. Para o conjunto de treinamento: classe 1: 53%, classe 0: 47%. Como

consequência, não foi necessário balancear as classes e, portanto, os resultados obtidos são os mais realistas. Além disso, os ajustes desse caso empregaram a menor quantidade de estimadores. O desempenho do modelo de melhor desempenho em cada caso é apresentado na Tabela 26.

Tabela 26 - Comparação entre modelos em diferentes limites de classificação

Nível de corte (ng. L ⁻¹)	Técnica	F1 - score macro averaged	Considerações
1.000	XGBoost	0,99	Alto desbalanceamento de classes de treinamento e teste, que foi corrigido pelo uso do SMOTE. Alto desempenho, mas inviável do ponto de vista prático.
500	XGBoost	0,90	O modelo foi capaz de classificar satisfatoriamente ambas as classes, mesmo com uma menor quantidade de amostras da classe minoritária (Preocupante, 0).
100	XGBoost	0,85	O modelo foi capaz de classificar satisfatoriamente ambas as classes, com equilíbrio no balanceamento entre as observações de treinamento e teste.

Para essa abordagem, os limites de corte de 500 ng. L⁻¹ e 100 ng. L⁻¹ foram os que melhor possibilitaram avaliar o potencial do uso de modelos de aprendizado na classificação de níveis de ocorrência de fármacos no efluente de ETE. Destacadamente, o limite restritivo de 100 ng. L⁻¹ se mostrou o mais viável para o monitoramento já que foi o que possibilitou o melhor balanceamento entre as classes.

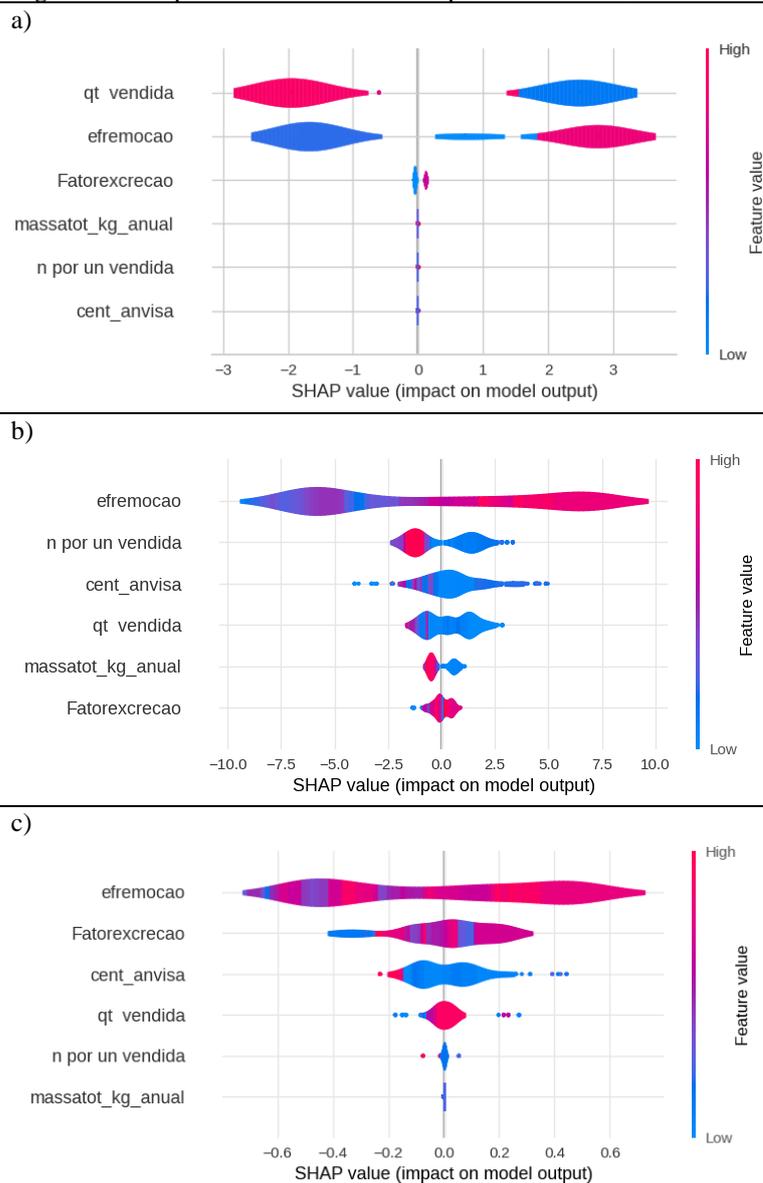
Assim, o limite de 1.000 ng. L⁻¹ não se mostrou vantajoso como um limite de monitoramento e, o limite intermediário de 500 ng. L⁻¹ e restritivo de 100 ng. L⁻¹, demonstraram – se mais adequados ao perfil de concentrações coletadas dos experimentos da literatura. Portanto, com base nos dados utilizados, esses limites proporcionam modelos de aprendizado com potencial de contribuir para o monitoramento da ocorrência de fármacos em ETE.

5.4.2.2 Importância das variáveis

Para todos os casos avaliados, os ajustes pelo algoritmo XGBoost apresentaram os maiores valores de métricas de avaliação em um tempo razoável para execução. Esse método provou ser o mais eficiente na separação das classes preditas, alcançando a menor quantidade de erros de classificação.

Para tanto, é importante compreender melhor quais critérios o algoritmo adotou para tomar suas decisões. Para os ajustes de melhor desempenho em cada caso, apresenta-se na Figura 299 a importância das variáveis e o impacto delas nas respostas obtidas.

Pela Figura 29(a) é possível notar que as variáveis “qt vendida” e “efremocao” foram as mais importantes para o modelo, dada a magnitude de seus impactos. Em seguida, “Fatorexcrecao” apresenta menor importância e impacto reduzido. As variáveis “massa_kg_anual”, “n por un vendida” e “cent_anvisa” não foram relevantes para o modelo de 1000 ng. L^{-1} . Observações com valores elevados de quantidade de medicamentos vendida apresentaram SHAP negativo, contribuindo para a ocorrência de concentração de saída preocupante. Baixos valores de eficiência de remoção contribuíram para a ocorrência de concentração preocupante, enquanto altos valores contribuíram para uma saída de concentração aceitável.

Figura 29 - Impacto das variáveis na resposta do modelo, Antibióticos

a) XGBoost – 1.000 ng. L⁻¹, b) XGBoost, 500 ng. L⁻¹, c) XGBoost, 100 ng. L⁻¹.

² Descrição das variáveis.

Fonte: Elaborado pela autora.

O Fator de Excreção apresenta um impacto próximo a zero, mas é possível verificar que valores reduzidos contribuíram para a ocorrência de concentração preocupante e altos valores contribuem para concentrações aceitáveis. Esse comportamento pode se justificar pelo fato de

² “qt vendida”: quantidade de embalagens comercializada anualmente; “efremocao”: Eficiência de Remoção de antibióticos no tratamento de ETE (%); “Fator Excreção”: Taxa de excreção do princípio ativo (%); “massa_tot_kg_anual”: massa total de fármaco comercializada anualmente (kg); “n por un vendida”: quantidade de unidades em uma embalagem comercializada; “cent_anvisa”: concentração anual de antibióticos comercializada (ng. L⁻¹).

que para este limite de corte alguns princípios ativos podem ter apresentado alta taxa de excreção e alta eficiência de remoção no tratamento, simultaneamente.

Para o ajuste de 500 ng. L⁻¹, Figura 29(b), as variáveis “efremocao”, “n por un vendida” e “cent_anvisa” apresentaram maior magnitude de impacto nas respostas do modelo. Verifica-se que altos valores de eficiência de remoção estão associados à ocorrência da classe aceitável, enquanto os valores medianos e reduzidos se associam à ocorrência da classe preocupante. Por outro lado, “n por un vendida” que representa a quantidade de unidades (comprimido, volume) presente em uma unidade comercializada (caixa, frasco) gerou um valor de SHAP negativo para valores elevados, contribuindo para uma concentração de saída preocupante. Quando essa variável apresentou baixo valor, contribuiu para a ocorrência da classe aceitável.

A concentração de entrada a partir das comercializações registradas na base da ANVISA (cent_anvisa) apresentou majoritariamente valores reduzidos, sendo possível identificar a ocorrência de valores altos associados à ocorrência de concentração preocupante. E os valores baixos se encontram associados à ocorrência de concentração aceitável.

A massa vendida anualmente (massa_kg_anual) e o Fator de excreção apresentaram baixo impacto no modelo. Ainda assim, é possível verificar que uma alta quantidade de massa comercializada contribuiu para a ocorrência da classe preocupante. O Fator de excreção apresentou valores altos, em sua maioria associados à ocorrência da classe aceitável e, pontualmente, valores baixos associados à concentração preocupante. Assim como no caso anterior, esse comportamento pode ser justificado, possivelmente, por observações de fármacos que mesmo com alto fator de excreção, apresentaram alta eficiência de remoção ou foram comercializados em baixa quantidade.

O modelo ajustado para o limite de corte de 100 ng. L⁻¹, Figura 29(c), é o mais restritivo para definir a concentração final como preocupante ou aceitável. A “efremocao”, “Fatorexcrecao” e “cent_anvisa” foram as variáveis de maior importância e com impacto de maior magnitude. Verifica-se que o impacto da Eficiência de remoção na ocorrência da classe preocupante (SHAP negativo) foi originado por um perfil mesclado entre valores baixos, medianos e altos. Entretanto, o impacto positivo foi afetado por altos valores da variável.

Assim, o lado associado ao SHAP negativo apresenta uma mistura de cores, que pode demonstrar que para esse limite restritivo de classificação, mesmo instâncias que apresentaram alta eficiência de remoção não foram capazes de atender a condição de classificação. Dessa forma, o lado associado à ocorrência de concentração preocupante engloba mais de um perfil de valores da variável, enquanto o lado associado a concentração aceitável nitidamente engloba apenas os valores altos de eficiência de remoção.

Algo semelhante ocorre para o Fator de excreção, visto que seus valores se associam à ocorrência de ambas as classes. No entanto, há maior incidência de observações associadas ao SHAP positivo, e, nessa situação, o impacto de maior magnitude se deu para observações com baixos valores de Fator de excreção. Isso indica que houve instâncias que apresentaram baixo fator de excreção, classificadas como “aceitável”, com alto impacto na classificação do modelo.

A concentração de entrada baseada nas comercializações da ANVISA apresentou menor impacto e, a ocorrência de concentração aceitável se associa a valores baixos dessa variável. Já a concentração preocupante se associa a valores reduzidos e elevados.

Por fim, a “quantidade vendida”, “n por un vendida” e “massa_kg_anual”, apresentaram baixíssimo impacto e não foram relevantes para a classificação nesse limite de corte.

5.5 ANÁLISE DE SENSIBILIDADE

Para cada caso investigado, foi realizada uma análise de sensibilidade do modelo de melhor desempenho, referido como modelo-base. A análise contou com a variação da vazão de esgoto afluente de cada município em três níveis distintos: 25% da vazão indicada nos dados do Atlas Esgotos 2019, 50% da vazão e 75% da vazão. Conseqüentemente, o valor da concentração predita de entrada na ETE (PEC_{inf}) para cada fármaco estudado também sofreu variação para cada nível.

Na Tabela 27 indica-se a vazão de esgoto considerada em cada nível para os municípios considerados.

Tabela 27 - Vazão de esgoto afluente (m^3) coletado e tratado para análise de sensibilidade

Município	Vazão – Atlas Esgotos	Nível 1 – Vazão 25%	Nível 2 – Vazão 50%	Nível 3 – Vazão 75%
Belo Horizonte	324216	81054	162108	243162
Campinas	120242,9	30060,72	60121,4	90182,16
Fortaleza	223862,4	55965,6	111931,2	167896,8
Jundiaí	79116,48	19779,12	39558,24	59337,36
Porto Alegre	181526,4	45381,6	90763,2	136144,8
Rio de Janeiro	817456,3	204364,1	408728,2	613092,2
São Carlos	47520	11880	23760	35640

Os valores indicados na Tabela são os mesmos para ambas as abordagens (Carbamazepina e Antibióticos). Entretanto, o valor de PEC_{inf} é específico para cada princípio ativo comercializado por município.

5.5.1 Carbamazepina

Os resultados da análise de sensibilidade para os modelos de Carbamazepina mostraram que independente do caso investigado, os resultados obtidos são semelhantes aos verificados quando se utiliza o valor total de vazão indicado pelo Atlas Esgotos no modelo base. Para todos os casos os modelos base foram XGboost. Os resultados seguem exibidos na Tabela 28.

Tabela 28 - Acurácia de modelos de Análise de Sensibilidade - Carbamazepina

Caso	Modelo base	Nível 1	Nível 2	Nível 3
1.000	0,80	0,85	0,85	0,85
500	0,94	0,94	0,94	0,94
100	0,94	0,79	0,79	0,79

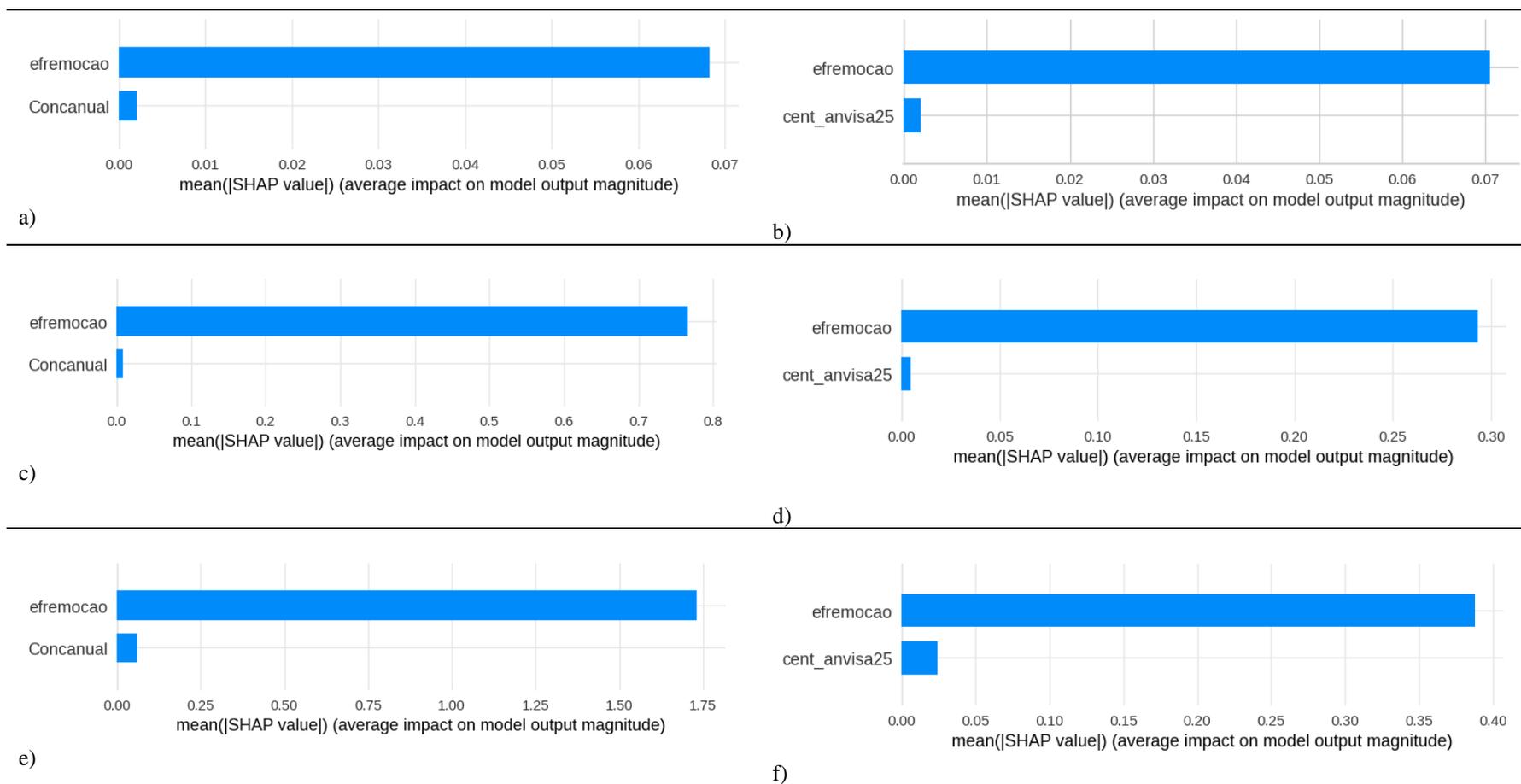
Para todos os níveis de sensibilidade os modelos atingiram a mesma acurácia. A acurácia obtida para o modelo base no caso de 1000 ng. L⁻¹ foi 80%, enquanto os modelos de sensibilidade apresentaram o valor de 85%. Para o caso de 500 ng. L⁻¹, o modelo base obteve 94% de acurácia e os modelos de sensibilidade alcançaram o mesmo desempenho.

O caso de 100 ng. L⁻¹ foi o que apresentou maior diferença entre a métrica avaliada, dado que o modelo base atingiu acurácia de 94% e os modelos de sensibilidade 79%. O resultado pode indicar que este modelo é menos robusto a variações na entrada.

Os valores de acurácia obtidos para os modelos dos casos de 1000 ng. L⁻¹ e 500 ng. L⁻¹ se encontram no Intervalo de Confiança ao nível de 95% calculado. Para os modelos base, IC_{95,1000} = [0.72; 0.88], IC_{95,500} = [0.89; 0.98], enquanto para o caso de 100 ng. L⁻¹ o IC_{95,100} = [0.89; 0.98], confirmando que esse limite demonstrou -se mais suscetível a variações.

Os resultados da Figura 3030 mostram o impacto médio das variáveis dos modelos base e modelos de sensibilidade.

Figura 30 - Impacto médio das variáveis, Análise de Sensibilidade de modelos de Carbamazepina



a) Caso 1.000 ng. L⁻¹, c) Caso 500 ng. L⁻¹, e) Caso 100 ng. L⁻¹, modelos base; b) Caso 1000 ng. L⁻¹, d) Caso 500 ng. L⁻¹, f) Caso 100 ng. L⁻¹, modelos para todos os níveis de sensibilidade em cada caso. *efremocao: Eficiência de Remoção de Carbamazepina (%); Concanual: Concentração anual de Carbamazepina comercializada (ng. L⁻¹).

Pela Figura 30 (a), (c), (e) é possível verificar que os modelos base de cada caso apresentaram o mesmo perfil de importância das variáveis. A “efremocao” foi o preditor mais importante e a “concanual” apresentou menor importância. Os modelos de sensibilidade da variável “concanual”, verificados na Figura 30 (b), (d) e (f) indicaram que a “efremocao” permaneceu sendo a variável de maior importância. Neste caso, “cent_anvisa25” indica o nível 1 de sensibilidade da “concanual” e sua imagem também representa os outros níveis de sensibilidade, já que os resultados foram idênticos.

Embora o caso de 100 ng. L⁻¹ tenha apresentado maiores diferenças no desempenho em comparação ao modelo base, ao se avaliar a importância das variáveis a importância da concentração de entrada proveniente dos dados da ANVISA é baixíssima, em ambos os modelos.

Dessa forma, os resultados da análise de sensibilidade apontaram que para a classificação da concentração final de Carbamazepina, a vazão de esgotos utilizada para o cálculo da concentração de entrada proveniente da comercialização de fármacos (PEC_{inf}) não exerceu influência significativa na saída do modelo, mesmo quando diferentes níveis de variação foram aplicados na entrada.

5.5.2 Antibióticos

Os resultados obtidos para a análise de sensibilidade dos casos de antibióticos podem ser verificados na Tabela 29.

Tabela 29 - Acurácia de modelos de Análise de Sensibilidade - Antibióticos

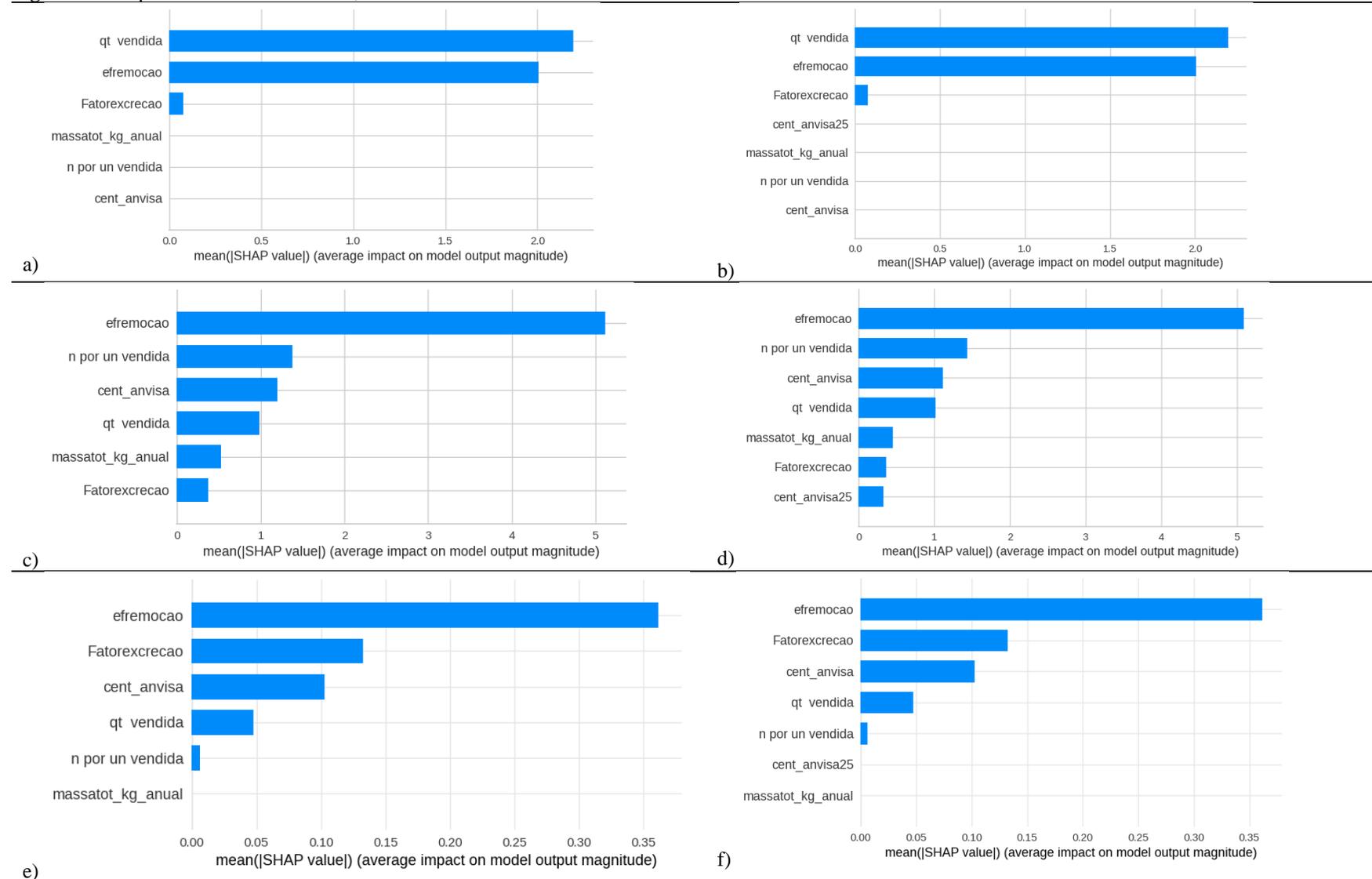
Caso	Modelo base	Nível 1	Nível 2	Nível 3
1.000	0,99	0,99	0,99	0,99
500	0,94	0,94	0,94	0,94
100	0,86	0,86	0,86	0,86

Para cada caso o modelo base utilizado foi o XGBoost, sendo que no caso de 1.000 ng. L⁻¹ foi obtida a maior acurácia (99%). Nesse caso, os modelos implementados para todos os níveis de sensibilidade também obtiveram acurácia de 99%. No caso de 500 ng. L⁻¹ tanto o modelo base como os modelos de sensibilidade apresentaram acurácia de 94%. E para o caso de 100 ng. L⁻¹, a acurácia do modelo base foi 86%, assim como a acurácia dos modelos de sensibilidade.

O Intervalo de Confiança da acurácia a 95% para o modelo base de cada caso foi: IC_{95,1000}[0,98; 1], IC_{95,500}[0,91; 0,97] e IC_{95,100}[0,80; 0,91]. Os modelos de sensibilidade

demonstraram resultados idênticos aos respectivos modelos base, sugerindo que os modelos base apresentaram resultados confiáveis e robustos uma vez que a variação na variável de entrada não alterou o desempenho obtido. Na Figura 3131 verifica-se o impacto das variáveis no modelo.

Figura 31 - Impacto médio das variáveis, Análise de Sensibilidade de modelos de Antibióticos.



a) Caso 1.000 ng. L⁻¹, b) Caso 500 ng. L⁻¹, c) Caso 100 ng. L⁻¹, modelos base; b) Caso 1000 ng. L⁻¹, d) Caso 500 ng. L⁻¹, f) Caso 100 ng. L⁻¹, modelos para todos os níveis de sensibilidade em cada caso. ³Descrição das variáveis.

³ “qt vendida”: quantidade de embalagens comercializada anualmente; “efremocao”: Eficiência de Remoção de antibióticos no tratamento de ETE (%); “Fator Excreção”: Taxa de excreção do princípio ativo (%); “massa_tot_kg_anual”: massa total de fármaco comercializada anualmente (kg); “n por un vendida”: quantidade de unidades em uma embalagem comercializada; “cent_anvisa”: concentração anual de antibióticos comercializada (ng. L⁻¹).

Para o modelo base do caso de 1.000 ng. L⁻¹, Figura 31 (a), apenas as variáveis “qt vendida”, “efremocao” e “Fatorexcrecao” apresentaram algum nível de impacto. Já para o caso de 500 ng. L⁻¹ todas as variáveis incluídas no modelo apresentaram importância. Para o modelo base do caso de 100 ng. L⁻¹, Figura 31 (e), a variável “massatotkganual” não apresentou impacto e “n_por_un_vendida” apresentou impacto baixíssimo.

Os resultados verificados para os modelos de sensibilidade, Figura 31 (b), (d) e (f) foram semelhantes aos modelos base, sendo que a variável de sensibilidade apresentou impacto baixo ou nulo.

Diante dos resultados verificados é possível afirmar que os modelos de sensibilidade atingiram alto nível de desempenho, assim como os modelos base. Além disso, não houve diferenças significativas em relação ao impacto das variáveis na resposta dos modelos mesmo ao se implementar variações na concentração de entrada.

Portanto, pode-se considerar que os modelos base construídos nesta pesquisa foram coerentes aos padrões indicados nos dados de entrada. Ainda que seja possível haver incertezas nos dados e estimativas coletadas do Atlas Esgotos, a análise de sensibilidade realizada demonstrou que os resultados dos modelos não sofrem alterações significativas.

5.6 RELEVÂNCIA DA ABORDAGEM NO CONTEXTO AMBIENTAL

O monitoramento dos CEC's é um desafio mundial, mesmo após décadas de pesquisas acerca dos impactos desses compostos para o meio ambiente e para a saúde humana. A limitação logística e financeira para a coleta de amostras, limitações analíticas, e variações nas características físico-químicas dos compostos se mostram como principais limitantes para o conhecimento pleno dos níveis de ocorrência e potenciais efeitos negativos associados a esses compostos.

As preocupações relacionadas à temática têm se agravado, especialmente, diante das mudanças climáticas e populacionais, dado que se espera maior frequência na ocorrência de eventos climáticos extremos como chuvas, inundações, secas e escassez de água, além do aumento populacional expressivo nas regiões em desenvolvimento, e, o aumento da taxa da população idosa, como discutido por Bunke *et al.* (2019).

Como resultado, espera-se o aumento do consumo de fármacos, a consequente disponibilização desses compostos nas matrizes ambientais e sua dispersão no ambiente. Nesse contexto, mitigar os efeitos negativos ocasionados por tal situação é urgente.

No Brasil, o cenário é ainda mais crítico visto que é um país de extensão continental, com intensas desigualdades sociais e, que ainda possui infraestrutura limitada para a prestação de serviços de saneamento. Nos registros publicados pela ANA (2017), foi estimado que apenas 42,6% do esgoto gerado era de fato coletado e tratado àquela época.

Nesse contexto, foi gerada uma carga de 9.098 t DBO/dia que, ao se considerar a fração de esgoto (I) não coletado, não tratado; (II) coletado, não tratado; (III) coletado, tratado; e, as soluções individuais (IV), estimou-se uma carga remanescente de 5.516 t DBO/dia lançada nos corpos hídricos receptores.

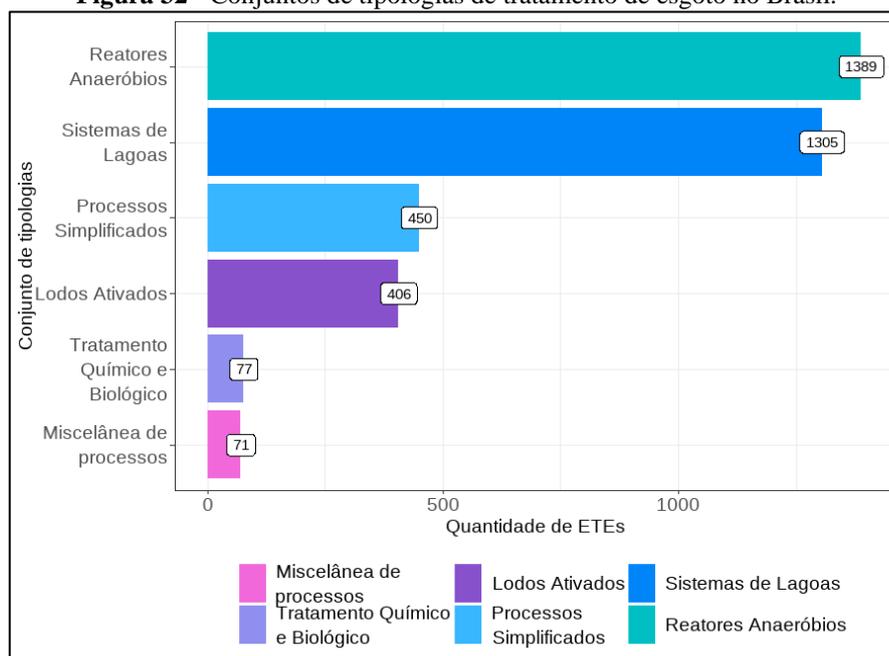
Dados publicados recentemente (MINISTÉRIO DAS CIDADES, 2017) apontaram que 52,2% do total de esgoto gerado nos municípios foi tratado para o ano-base de 2022. Esse aumento verificado para os últimos anos não anula o fato de que a taxa da população brasileira com acesso à esgoto tratado (56%) é notoriamente inferior à de países como os Países Baixos, Noruega, Japão e Irlanda, onde 99,7%, 87,5%, 81% e 66,1% da população encontra-se conectada à alguma ETE, segundo dados do ano de 2022 (ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2024).

Essas estatísticas reforçam que para a realidade brasileira, o desafio para ampliação do acesso ao esgoto tratado depende, preliminarmente, da universalização da prestação de serviços de coleta de esgoto, além do aumento do percentual de tratamento de esgoto coletado, como definido pelas metas do Plansab (BRASIL, 2014).

O país conta com 3.668 ETE que podem atender um ou mais municípios e os principais conjuntos de tipologias de tratamento são Reatores Anaeróbios (37%), seguido por Sistemas de Lagoas de Estabilização (35%). A Figura 322 ilustra a distribuição das tipologias de tratamento verificadas nas ETEs constatadas pelo monitoramento (ANA, 2020).

As tipologias empregadas compreendem diferentes possibilidades de arranjos tecnológicos e se caracterizam pela faixa de eficiência de remoção de matéria orgânica, em termos de Demanda Bioquímica de Oxigênio (DBO) e, de modo complementar, também podem apresentar algum nível de remoção de patógenos e nutrientes. Entretanto, é comum que a eficiência dos tratamentos seja afetada por problemas operacionais que levem à resultados abaixo do esperado.

Esse panorama mostra que se ainda existe uma limitação para a operação plena das ETEs em relação a parâmetros físico-químicos cujos tratamentos já se encontram bem fundamentados, o estabelecimento de rotinas para a remoção de CPEs sequer é uma pauta em planejamento.

Figura 32 - Conjuntos de tipologias de tratamento de esgoto no Brasil.

Fonte: Elaborado pela autora a partir de dados do Atlas Esgotos (ANA, 2017).

Para contextos como esses, recorrentes em países em desenvolvimento, Acuña *et al.* (2020) avaliaram que apenas a melhoria de tecnologias seria insuficiente para mitigar a poluição ambiental devido à disponibilização indevida dos fármacos no meio ambiente. Os autores consideraram o Diclofenaco na pesquisa, e identificaram que unicamente a implementação do ODS 6 considerando a redução pela metade da população sem acesso à saneamento e a melhoria dos efluentes das estações de tratamento seria a medida mais efetiva, atingindo valores inferiores a 100 ng. L⁻¹ do IFA em águas superficiais.

Assim, a carência na infraestrutura de saneamento se sobressai como aspecto determinante e, como depende de recursos financeiros, verifica-se a confluência entre aspectos socioeconômicos e saúde pública impactando a poluição ambiental por fármacos.

Por isso, se tornou inadiável investigar, conceber e implementar tecnologias com capacidade de assistir o gerenciamento em países em desenvolvimento, já que esses locais são os mais suscetíveis à baixa cobertura de tratamento de esgoto e, portanto, a ocorrência de elevadas concentrações de fármacos, como verificado por Wilkinson *et al.* (2022).

O avanço tecnológico é fundamental para auxiliar o monitoramento dos fármacos, especialmente porque os dados de ocorrência são limitados e obtê-los envolve um processo custoso, que por vezes, depende da aceitação das empresas de saneamento para a realização de análises e utilização de seus dados. Devido a isso, a modelagem se configura como uma ferramenta importante na estimativa dos níveis de ocorrência dos fármacos, possibilitando combinar dados de comercialização e dados da remoção em ETEs previamente monitoradas.

Os modelos baseados em dados contribuíram para a compreensão de padrões de ocorrência desses compostos nos efluentes das ETEs, por meio da capacidade de aprendizado. Assim, considerou-se nesta pesquisa a avaliação do potencial dessas técnicas para o problema investigado e, ainda, foi selecionada uma abordagem próxima do processo de decisão humana, a classificação.

Os modelos desenvolvidos nesta pesquisa partiram da consideração de que todo o conteúdo de fármaco contido em uma unidade comercializada foi consumido pelo paciente, assim como apontado por He *et al.* (2020). Portanto, não se considerou o consumo parcial ou outras formas de aquisição além das farmácias e drogarias. Essa ponderação é relevante uma vez que é um aspecto chave que pode influenciar nas estimativas de consumo e posterior predição da concentração excretada.

Os dados da remoção de fármacos em ETE foram coletados de diferentes estudos de caso da literatura. Com base nos critérios expostos no item 4.2.4, os dados publicados foram selecionados para ofertar aos modelos um padrão de aprendizado. Porém, é importante refletir sobre potenciais incertezas nas medições devido à fatores operacionais e analíticos, além da alta variabilidade encontrada.

Como discutido, o monitoramento *in loco* é demasiadamente custoso. Assim, uma opção foi justamente coletar dados de localidades onde já havia sido feito algum monitoramento. Adicionalmente, os limites de rotulação para a classificação correspondem a níveis regulatórios para a ocorrência de antibióticos e Carbamazepina em águas superficiais, já que diante da escassez de regulamentações no tema, não foram encontrados níveis de fármacos recomendados para o padrão de lançamento de efluentes sanitários.

Mesmo diante das delimitações consideradas, esta pesquisa demonstrou que as técnicas de aprendizado de máquina foram capazes de gerar modelos de alto desempenho cujo aprendizado refletiu os padrões verificados nas variáveis preditoras. Para a classificação de concentrações de Carbamazepina os modelos apresentaram alto desempenho, mas foram impactados apenas pela Eficiência de Remoção. Esse comportamento afeta o potencial dessa abordagem, pois invalida os efeitos da comercialização do fármaco para a construção do modelo de classificação.

Nesse caso, foi verificado que a concentração afluente proveniente da comercialização do fármaco não foi um preditor relevante dada a complexidade dos mecanismos que afetam sua disponibilização ambiental. Nesse sentido, os resultados indicaram que um modelo representativo para a Carbamazepina necessita da inclusão de novas variáveis, primordialmente, parâmetros operacionais em conjunto com propriedades físico-químicas.

Já o mecanismo de modelagem proposto para os Antibióticos demonstrou ser adequado para classificar os níveis de ocorrência e, as respostas dos modelos foram significativamente influenciadas pela Eficiência de remoção, Fator de Excreção e a Concentração de entrada de fármacos. O potencial dessa abordagem foi evidenciado tanto pelas métricas de desempenho quanto pela interpretação do modelo, validando o impacto da concentração de entrada proveniente da comercialização dos fármacos no modelo.

Na literatura científica, as aplicações de métodos de aprendizado de máquina para a predição de fármacos em efluentes e águas são bastante restritas e recentes (TAOUFIK et al., 2022). Especificamente em relação aos métodos *boosting*, observa-se que têm atingido alta capacidade preditiva tanto em abordagens de classificação como de regressão (AHMED et al., 2024; CHEN et al., 2025; NGUYEN et al., 2024; YAQUB et al., 2022). Isso mostra que são capazes de identificar de forma mais assertiva os fatores que afetam o mecanismo de degradação dos compostos. No entanto, os fármacos são selecionados com base em critérios específicos e, portanto, os ingredientes ativos e variáveis consideradas diferem entre os estudos.

Os resultados obtidos nesta pesquisa reforçam uma nova possibilidade de modelagem com o potencial de auxiliar gestores ambientais a monitorar os níveis de fármacos prioritários lançados em corpos hídricos de municípios brasileiros. Nesse mecanismo, apenas a consideração dos dados de consumo e registros prévios da remoção de fármacos em ETE se mostrou suficiente para obter uma estimativa assertiva, especialmente para os antibióticos. Para a Carbamazepina, sugere-se a complementação com novas variáveis.

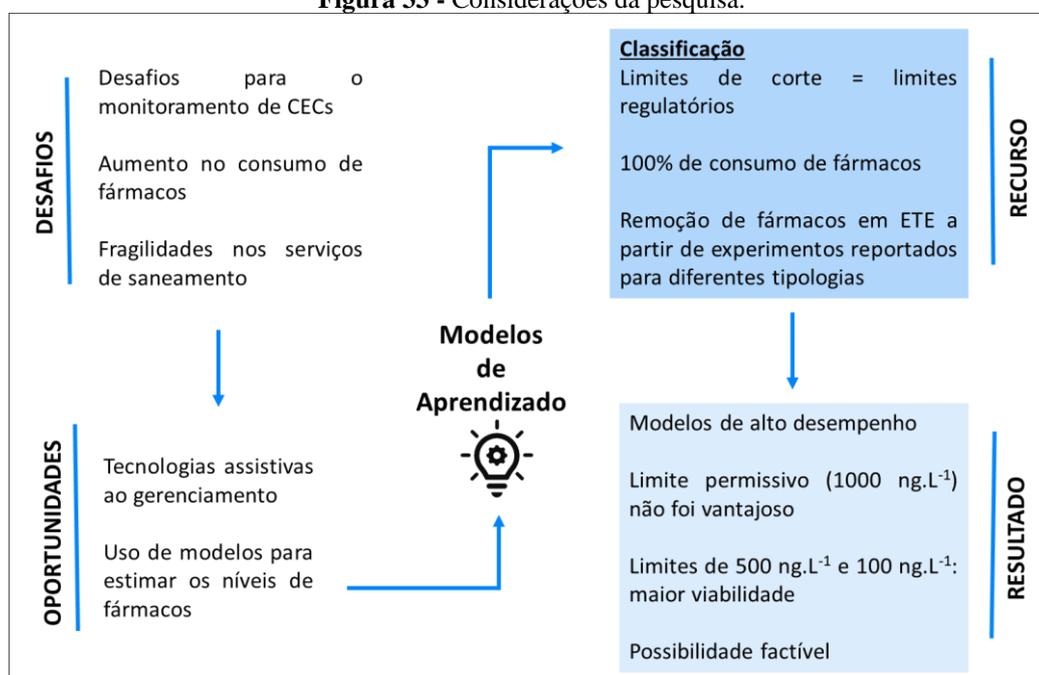
Os resultados verificados são particularmente importantes para o contexto no qual não há possibilidade de se realizar campanhas amostrais para quantificar a ocorrência dos fármacos no meio ambiente. Assim, a identificação de fármacos prioritários ao monitoramento e o conhecimento dos fatores mais influentes para a ocorrência ambiental pode direcionar estratégias de sensibilização para a redução do consumo exacerbado de fármacos e melhoramento do processo de tratamento em ETE.

Além disso, foi possível avaliar como diferentes limites regulatórios afetariam o monitoramento desses compostos. Nesse caso, os limites 500 ng. L⁻¹ e 100 ng. L⁻¹ seriam os mais adequados para uma possível implementação, já que responderam bem às concentrações identificadas em campo. No entanto, avaliações devem ser realizadas a fim de determinar se esses são limites possíveis de serem atingidos frente à capacidade das ETE do Brasil.

Os modelos confirmaram que técnicas de Aprendizado de Máquina, em especial modelos de classificação baseados em árvores com abordagem *boosting*, podem ser uma ferramenta factível para auxiliar a estimativa dos níveis de ocorrência de fármacos, reduzindo

a exigência de amplas coletas de dados. Além disso, os limites de rotulação utilizados simularam como diferentes limites regulatórios podem interferir a resposta de cada modelo. O esquema apresentado na Figura 33 compila as principais considerações realizadas.

Figura 33 - Considerações da pesquisa.



Fonte: Elaborado pela autora.

Neste sentido, os resultados demonstraram que é possível estabelecer um mecanismo para predição da concentração de fármacos destinados aos cursos d'água receptores de esgotos tratados por meio de técnicas de aprendizado de máquinas, de forma a viabilizar a possibilidade de monitoramento desses compostos, especialmente em localidades onde há empecilhos para o monitoramento *in loco*.

6 CONCLUSÕES

A pesquisa sobre o desenvolvimento da proposta de um mecanismo de classificação capaz de prever o nível das concentrações de alguns fármacos nos efluentes de estações de tratamento de esgoto (ETE) permitiu concluir que:

- Os fármacos Ampicilina, Amoxicilina, Azitromicina, Cefalexina, Ciprofloxacino, Claritromicina, Norfloxacino, Ofloxacino e Carbamazepina foram priorizados como os de maior relevância presentes nos esgotos;
- De modo geral, a Eficiência de remoção dos fármacos nas ETEs apresentou alta variabilidade. Os antibióticos apresentaram eficiência de remoção intermediária (mediana 55%) as drogas psiquiátricas demonstraram ser mais resistentes ao tratamento (mediana = 24%) e, os anti-inflamatórios apresentaram maior potencial de remoção (mediana = 74%);
- Os modelos de classificação propostos obtiveram desempenho satisfatório e evidenciaram que a escolha dos limites de rotulação para as classes de concentração de saída baseada em valores normativos influenciou significativamente as repostas dos modelos;
- O limite de rotulação permissivo (1000 ng. L^{-1}) foi o mais desafiador devido à distribuição das classes, enquanto os limites de 500 ng. L^{-1} e 100 ng. L^{-1} demonstraram – se mais viáveis em relação ao desempenho e interpretação.
- Para todos os casos, ajustes XGBoost superaram o Adaboost e Gradient Boosting e, a interpretação dos modelos se mostrou condizente ao esperado. Os modelos de antibióticos demonstraram maior potencial do que os de Carbamazepina, já que nas decisões, consideraram a concentração de entrada de fármacos proveniente da comercialização.

Nesse sentido, a elaboração dos modelos auxiliou a avaliação do potencial de uso de uma abordagem não convencional baseada em Aprendizado de Máquina, por métodos de classificação Boosting, como ferramenta de apoio ao monitoramento da ocorrência de fármacos em território brasileiro, considerando registros de venda desses compostos como variáveis preditoras.

A quantificação das incertezas das bases de dados utilizadas nesta pesquisa não foi realizada. Embora a metodologia desenvolvida tenha sido rigorosa, considera-se que esse é um fator essencial para apoiar as evidências obtidas em estudos originados a partir desses dados, o que evidencia a necessidade desse tópico ser investigado em futuras pesquisas.

Assim, os modelos construídos são importantes alternativas complementares ao monitoramento e controle de fármacos no esgoto sanitário, auxiliando na promoção da saúde pública e ainda, podem eximir a necessidade de realizar campanhas amostrais frequentes para quantificação da concentração desses componentes. A abordagem tem potencial para ser adaptada para outras localidades.

7 SUGESTÕES PARA TRABALHOS FUTUROS

- Construir Modelos de aprendizado para predição da ocorrência de fármacos considerando aspectos temporais, como estações do ano;
- Investigar o perfil de comercialização de fármacos em municípios brasileiros considerando informações sociodemográficas;
- Inserir parâmetros de processo de ETE, novas variáveis provenientes da comercialização e, considerar outros psicotrópicos para avaliar de forma precisa os mecanismos de classificação para essa classe terapêutica. Sugere-se buscar dados de ocorrência de Clonazepam e Fluoxetina;
- Avaliar aspectos relacionados ao *deploy* do mecanismo desenvolvido;
- Complementar base de dados de ocorrência de fármacos em ETE com novos dados, atualizados em buscas recentes na literatura e estruturar base de dados em painel online;
- Investigar o mecanismo proposto considerando municípios que não tenham situação estabelecida para coleta e tratamento de esgotos (Grupo B do Atlas Esgotos), a partir de valores de vazão de esgoto baseados nas diferentes situações de prestação do serviço indicados na base disponibilizada pela Agência Nacional de Águas e Saneamento Básico;
- Avaliar o potencial de modelos de *Quantum Machine Learning* para a classificação de níveis de ocorrência de fármacos frente à inclusão de novas variáveis (química quântica e molecular), velocidade de execução e acurácia dos modelos;
- Quantificar as incertezas da base de dados de dispensação de fármacos, do SNGPC, disponibilizada pela ANVISA

REFERÊNCIAS

ACOSTA-ANGULO, B. et al. Unveiling the potential of machine learning in cost-effective degradation of pharmaceutically active compounds: A stirred photo-reactor study. **Chemosphere**, v. 358, p. 142222, jun. 2024.

ACUÑA, V. et al. Management actions to mitigate the occurrence of pharmaceuticals in river networks in a global change context. **Environment International**, v. 143, p. 105993, out. 2020.

AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO (ANA); SECRETARIA NACIONAL DE SANEAMENTO AMBIENTAL (SNSA). **Atlas esgotos : despolição de bacias hidrográficas**. [s.l.] Brasília: ANA, 2017., 2017.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. Resolução RDC N° 57, de 17 de novembro de 2009. Dispõe sobre o registro de insumos farmacêuticos ativos (IFA) e dá outras providências. **Diário Oficial União**, 2009.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. **Venda de Medicamentos Controlados e Antimicrobianos - Medicamentos Industrializados**. Disponível em: <<https://dados.gov.br/dados/conjuntos-dados/venda-de-medicamentos-controlados-e-antimicrobianos---medicamentos-industrializados>>. Acesso em: 16 set. 2024.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. Resolução RDC n° 471, de 23 de fevereiro de 2021. Dispõe sobre os critérios para a prescrição, dispensação, controle, embalagem e rotulagem de medicamentos à base de substâncias classificadas como antimicrobianos de uso sob prescrição, isoladas ou em associação, listadas em Instrução Normativa específica. **Diário Oficial da União**, 2021.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA; SECRETARIA DE VIGILÂNCIA EM SAÚDE; MINISTÉRIO DA SAÚDE. Portaria N° 344, de 12 de maio de 1998. Aprova o Regulamento Técnico sobre substâncias e medicamentos sujeitos a controle especial. **Diário Oficial União**, 1998.

AHMED, Y. et al. A novel interpretable machine learning and metaheuristic-based protocol to predict and optimize ciprofloxacin antibiotic adsorption with nano-adsorbent. **Journal of Environmental Management**, v. 370, p. 122614, nov. 2024.

ANDERSON, J. C. et al. Evaluation of cold-weather wastewater nitrification technology for removal of polar chemicals of emerging concern from rural Manitoba wastewaters. **Chemosphere**, v. 253, p. 126711, 2020.

ANVISA. Resolução - RDC N° 44, de 26 de Outubro de 2010. out. 2010.

ANVISA. Resolução – RDC N° 20, de 5 de maio de 2011. Dispõe sobre o controle de medicamentos à base de substâncias classificadas como antimicrobianos, de uso sob prescrição, isoladas ou em associação. **Diário Oficial União**, 2011.

ANVISA. Resolução nº 174, de 15 de setembro de 2017. Dispõe sobre a atualização da lista de antimicrobianos registrados na Anvisa. **Diário Oficial da União**, 2017.

ANVISA. **Anuário Estatístico do Mercado Farmacêutico 2017**. Anvisa Brasília, DF, , 2018.

ANVISA. **Anuário Estatístico do Mercado Farmacêutico - 2022**. Brasília: Anvisa, 2023, , 2023. Disponível em: <<https://www.gov.br/anvisa/pt-br/centraisdeconteudo/publicacoes/medicamentos/cmed/anuario-estatistico-2022/view>>. Acesso em: 16 set. 2024

ANVISA; MINISTÉRIO DA SAÚDE. Resolução da Diretoria Colegiada - RDC N° 68, de 28 de novembro de 2014. Dispõe sobre a atualização do Anexo I, Lista de Antimicrobianos Registrados na Anvisa, da Resolução – RDC nº 20, de 5 de maio de 2011 e dá outras providências. **Diário Oficial União**, 2014.

ASSOCIAÇÃO DOS CONSUMIDORES. Relatório Anual 2015. **Porteste. A nossa voz impõe respeito.**, p. 40, 2015.

BACK, J. T.; ASSINI, F. Análise das prescrições de psicotrópicos em farmácias privadas na cidade de Monte Carlo, Santa Catarina. **Revista Eletrônica de Farmácia**, v. 14, n. 2, 18 set. 2017.

BAEK, S.-S. et al. Replacing the internal standard to estimate micropollutants using deep and machine learning. **Water Research**, v. 188, p. 116535, 2021.

BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. [s.l.] Springer, 2006. v. 4

BRASIL. **Plano Nacional de Saneamento Básico (PLANSAB)**. Brasília, DF: Minsitério das Cidades, 2014.

BRASIL, C. Resolução no 430, de 13 de maio de 2011. Em: **Dispõe sobre as condições e padrões de lançamento de efluentes, complementa e altera a Resolução**. [s.l.: s.n.]. v. 357.

CAETANO, M. C. et al. Consumo de antimicrobianos nas farmácias e drogarias privadas brasileiras à luz do PAN-BR e da pandemia de COVID-19. **Brazilian Journal of Development**, v. 8, n. 1, p. 645–669, 2022.

CARRASCO, B. D.; BARROS, R. R. Consumo nacional de antimicrobianos prescritos por cirurgiões-dentistas. **Rev. Flum. Odontol.(Online)**, p. 110–120, 2024.

CASAS, A. et al. Machine learning screening tools for the prediction of extraction yields of pharmaceutical compounds from wastewaters. **Journal of Water Process Engineering**, v. 62, p. 105379, maio 2024.

CASTIGLIONI, S. et al. Mass balance of emerging contaminants in the water cycle of a highly urbanized and industrialized area of Italy. **Water Research**, v. 131, p. 287–298, mar. 2018.

CHAWLA, N. V. et al. SMOTE: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, v. 16, p. 321–357, 2002.

CHEMSPIDER. **ChemSpider: Search and Share Chemistry - Homepage**.

Disponível em: <<https://www.chemspider.com/>>. Acesso em: 13 out. 2024.

CHEN, L. et al. Emerging organic contaminants and odorous compounds in secondary effluent wastewater: Identification and advanced treatment. **Journal of Hazardous Materials**, v. 408, p. 124817, abr. 2021.

CHEN, L. et al. Quality evaluation parameter and classification model for effluents of wastewater treatment plant based on machine learning. **Water Research**, v. 268, p. 122696, jan. 2025.

CHEN, T.; GUESTRIN, C. **XGBoost: A Scalable Tree Boosting System**.

Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. **Anais...: KDD '16**. New York, NY, USA: Association for

Computing Machinery, 2016. Disponível em:

<<https://dl.acm.org/doi/10.1145/2939672.2939785>>. Acesso em: 16 set. 2024

COUTO, C. F.; LANGE, L. C.; AMARAL, M. C. S. Occurrence, fate and removal of pharmaceutically active compounds (PhACs) in water and wastewater treatment plants—A review. **Journal of Water Process Engineering**, v. 32, p. 100927, 2019.

DE OLIVEIRA SANTOS, D.; BUENO, L. S.; DE PASSOS, S. G. Uso indiscriminado de psicotrópicos e o aumento das emergências psiquiátricas. **Revista JRG de Estudos Acadêmicos**, v. 7, n. 14, p. e141164–e141164, 2024.

DE VOOGT, P. et al. Development of a common priority list of pharmaceuticals relevant for the water cycle. **Water Science and Technology**, v. 59, n. 1, p. 39–46, 2009.

DEL FIOL, F. DE S. et al. Consumption trends of antibiotics in Brazil during the COVID-19 pandemic. **Frontiers in Pharmacology**, v. 13, p. 844818, 2022.

DOLATABADI, M. et al. Adsorption of tetracycline antibiotic onto modified zeolite: Experimental investigation and modeling. **MethodsX**, v. 7, p. 100885, 2020.

DRUGBANK. **Database for Drug and Drug Target Info**. Disponível em: <<https://go.drugbank.com/>>. Acesso em: 13 out. 2024.

DUARTE, M. S. et al. A Review of Computational Modeling in Wastewater Treatment Processes. **ACS ES&T Water**, v. 4, n. 3, p. 784–804, 8 mar. 2024.

EPA. **AQUATIC LIFE CRITERIA FOR CONTAMINANTS OF EMERGING CONCERN PART I GENERAL CHALLENGES AND RECOMMENDATIONS**. 2008. Disponível em: <https://www.epa.gov/sites/default/files/2015-08/documents/white_paper_aquatic_life_criteria_for_contaminants_of_emerging_concern_part_i_general_challenges_and_recommendations_1.pdf>.

EPA. **United States Environmental Protection Agency (EPA)**. Disponível em: <<https://www.epa.gov/ccl>>. Acesso em: 23 ago. 2023.

EUROPEAN COMMISSION. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. **Official Journal of the European Communities**, v. 37, p. 1–72, 2000.

EUROPEAN COMMISSION. Commission Implementing Decision (EU) 2015/495 of 20 March 2015 establishing a watch list of substances for Union-wide monitoring in the field of water policy pursuant to Directive 2008/105/EC of the European Parliament and of the Council. **Off. J. Eur. Union L**, v. 78, p. 40–42, 2015.

EUROPEAN COMMISSION. Commission implementing decision (EU) 2018/840 of 5 June 2018 establishing a watch list of substances for union-wide monitoring in the field of water policy pursuant to directive 2008/105/EC of the European Parliament and of the council and repealing commission implementing decision (EU) 2015/495 (notified under document C(2018) 3362). **Official Journal of the European Union**, p. 9–12, 2018.

EUROPEAN COMMISSION. Commission Implementing Decision (EU) 2020/1161 of 4 August 2020 establishing a watch list of substances for Union-wide monitoring in the field of water policy pursuant to Directive 2008/105/EC of the European Parliament and of the Council (notified under document number C(2020) 5205) (Text with EEA relevance). **Official Journal of the European Union**, 2020.

EUROPEAN MEDICINES AGENCY (EMA); COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (CHMP). Environmental risk assessment of medicinal products for human use. **Scientific guideline**, 2006.

FAN, M. et al. A review on experimental design for pollutants removal in water treatment with the aid of artificial intelligence. **Chemosphere**, v. 200, p. 330–343, jun. 2018.

FELICIANO, A. R. S. et al. Advanced oxidation processes employment for the degradation of lamivudine: kinetic assessment, toxicity study and mathematical modeling. **Water Quality Research Journal**, v. 55, n. 3, p. 249–260, 2020.

FONSECA, N. et al. Assessment of Non-Steroidal Anti-Inflammatories in River Waters of Northeastern of Brazil: Occurrence and Environmental Risk. **Journal of the Brazilian Chemical Society**, 2024.

FOSSO-KANKEU, E. (ED.). **Nano and bio-based technologies for wastewater treatment: prediction and control tools for the dispersion of pollutants in the environment**. Beverly, MA : Hoboken, NJ: Scrivener Publishing ; Wiley, 2019.

FRIEDMAN, J. H. 1999 Reitz lecture Greedy function approximation: A Gradient Boosting Machine. **The Annals of Statistics**, v. 29, n. 5, p. 1189–1232, 2001.

GEISSEN, V. et al. Emerging pollutants in the environment: A challenge for water resource management. **International Soil and Water Conservation Research**, v. 3, n. 1, p. 57–65, mar. 2015.

GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. [s.l.] O'Reilly Media, Inc., 2022.

GONZÁLEZ PEÑA, O. I.; LÓPEZ ZAVALA, M. Á.; CABRAL RUELAS, H. Pharmaceuticals market, consumption trends and disease incidence are not driving the pharmaceutical research on water and wastewater. **International journal of environmental research and public health**, v. 18, n. 5, p. 2532, 2021.

GROSSELI, G. M. **Contaminantes emergentes em estações de tratamento de esgoto aeróbia e anaeróbia**. Tese—São Carlos - SP: Universidade Federal de São Carlos (Departamento de Química Programa de Pós-Graduação em Química), 2016.

GROSS-SOROKIN, M. Y.; ROAST, S. D.; BRIGHTY, G. C. Assessment of Feminization of Male Fish in English Rivers by the Environment Agency of England and Wales. **Environmental Health Perspectives**, v. 114, n. Suppl 1, p. 147–151, abr. 2006.

HADJIMICHAEL, A.; COMAS, J.; COROMINAS, L. Do machine learning methods used in data mining enhance the potential of decision support systems? A review for the urban water sector. **AI Communications**, v. 29, n. 6, p. 747–756, 2016.

HAMADA, M. S.; ZAQOOT, H. A.; SETHAR, W. A. Using a supervised machine learning approach to predict water quality at the Gaza wastewater treatment plant. **Environmental Science: Advances**, v. 3, n. 1, p. 132–144, 2024.

HASTIE, T.; FRIEDMAN, J.; TIBSHIRANI, R. **The Elements of Statistical Learning**. 2 ed. ed. New York, NY: Springer New York, 2001.

HE, K. et al. Sale-based estimation of pharmaceutical concentrations and associated environmental risk in the Japanese wastewater system. **Environment International**, v. 139, p. 105690, 1 jun. 2020.

HE, L. et al. Applications of computational chemistry, artificial intelligence, and machine learning in aquatic chemistry research. **Chemical Engineering Journal**, v. 426, p. 131810, dez. 2021.

HOEFLER, R. et al. Trends in sales of antidepressants in Brazil from 2014 to 2020: A time trend analysis with joinpoint regression. **Journal of Affective Disorders**, v. 323, p. 213–218, fev. 2023.

INSTITUTE FOR HUMAN DATA SCIENCE (IQVIA). Global Use of Medicines 2023: Outlook 2027. 2023.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **PNSB - Pesquisa Nacional de Saneamento Básico**. Disponível em: <<https://www.ibge.gov.br/estatisticas/multidominio/meio-ambiente/9073-pesquisa-nacional-de-saneamento-basico.html?edicao=18098&t=sobre>>.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **PNAD - Pesquisa Nacional por Amostra de Domicílios**. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/populacao/9127-pesquisa-nacional-por-amostra-de-domicilios.html>>.

JAMES, G. et al. **An Introduction to Statistical Learning: with Applications in Python**. Cham: Springer International Publishing, 2023.

JIANG, J. et al. Optimization of a Novel Engineered Ecosystem Integrating Carbon, Nitrogen, Phosphorus, and Sulfur Biotransformation for Saline Wastewater Treatment Using an Interpretable Machine Learning Approach. **Environmental Science & Technology**, v. 58, n. 29, p. 12989–12999, 23 jul. 2024.

KOSTICH, M. S.; BATT, A. L.; LAZORCHAK, J. M. Concentrations of prioritized pharmaceuticals in effluents from 50 large wastewater treatment plants in the US and implications for risk estimation. **Environmental pollution**, v. 184, p. 354–359, 2014.

KÖTKE, D. et al. Occurrence and environmental risk assessment of pharmaceuticals in the Mondego river (Portugal). **Heliyon**, v. 10, n. 15, p. e34825, ago. 2024.

KOT-WASIK, A.; JAKIMSKA, A.; ŚLIWKA-KASZYŃSKA, M. Occurrence and seasonal variations of 25 pharmaceutical residues in wastewater and drinking water

treatment plants. **Environmental Monitoring and Assessment**, v. 188, n. 12, p. 661, dez. 2016.

LEMAĂŽTRE, G.; NOGUEIRA, F.; ARIDAS, C. K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. **Journal of machine learning research**, v. 18, n. 17, p. 1–5, 2017.

LI, R. et al. Enhanced Insights into Effluent Prediction in Wastewater Treatment Plants: Comprehensive Deep Learning Model Explanation Based on SHAP. **ACS ES&T Water**, v. 4, n. 4, p. 1904–1915, 12 abr. 2024.

LI, Y. et al. Distribution, Removal, and Risk Assessment of Pharmaceuticals and Their Metabolites in Five Sewage Plants. **International Journal of Environmental Research and Public Health**, v. 16, n. 23, p. 4729, 27 nov. 2019.

LIMA, D. R. S. et al. Fármacos e desreguladores endócrinos em águas brasileiras: ocorrência e técnicas de remoção. **Engenharia Sanitaria e Ambiental**, v. 22, n. 6, p. 1043–1054, dez. 2017a.

LIMA, M. A. D. M. et al. Uso do Atlas Esgotos: Despoluição de Bacias Hidrográficas no desenvolvimento de trabalhos científicos. **Revista Sustinere**, v. 8, n. 1, 10 jul. 2020.

LIMA, M. G. et al. Indicators related to the rational use of medicines and its associated factors. **Revista de Saúde Pública**, v. 51, p. 23s, 22 set. 2017b.

LOPES, L. C.; MOTTER, F. R.; CARVALHO-SOARES, M. D. L. Consumption of antibiotics in Brazil - an analysis of sales data between 2014 and 2019. **Antimicrobial Resistance & Infection Control**, v. 13, n. 1, p. 60, 9 jun. 2024.

LOPES, V. S. A. et al. Development of a solid-phase extraction system modified for preconcentration of emerging contaminants in large sample volumes from rivers of the lagoon system in the city of Rio de Janeiro, Brazil. **Marine Pollution Bulletin**, v. 110, n. 1, p. 572–577, set. 2016.

LUBISCO, N. M. L.; VIEIRA, S. C. **Manual de estilo acadêmico: trabalhos de conclusão de curso, dissertações e teses**. Salvador: Edufba, 2019.

LUNDBERG, S. A unified approach to interpreting model predictions. **arXiv preprint arXiv:1705.07874**, 2017.

MELÉNDEZ-MARMOLEJO, J. et al. Contaminantes emergentes. Problemática ambiental asociada al uso de antibióticos. Nuevas técnicas de detección, remediación y perspectivas de legislación en América Latina. **Revista de salud ambiental**, v. 20, n. 1, p. 53–61, 2020.

MENDES, C. et al. Modeling of simultaneous carbon and nitrogen removal (SCNR) in an anaerobic/anoxic reactor treating salmon fishery wastewater. **Journal of Water Process Engineering**, v. 56, p. 104442, dez. 2023.

METCALF & EDDY. **Wastewater engineering: treatment and resource recovery**. 5a. ed. New York, NY: McGraw-Hill Education, 2014.

MICHIE, D. et al. (EDS.). **Machine learning, neural and statistical classification**. 1. publ., 2. [print.] ed. New York: Ellis Horwood, 1995.

MIKULIC. **Top national pharma markets by market share 2023**. Disponível em: <<https://www.statista.com/statistics/245473/market-share-of-the-leading-10-global-pharmaceutical-markets/>>. Acesso em: 13 out. 2024.

MIKULIC, M. **Pharmaceutical market worldwide revenue 2001-2023**. Disponível em: <<https://www.statista.com/statistics/263102/pharmaceutical-market-worldwide-revenue-since-2001/>>. Acesso em: 16 set. 2024.

MINISTÉRIO DAS CIDADES. **SNIS**. Disponível em: <<https://www.gov.br/cidades/pt-br/acao-a-informacao/acoes-e-programas/saneamento/snis>>. Acesso em: 13 out. 2024.

MOLINA, R.; DE LA BLANCA, N. P.; TAYLOR, C. C. Modern statistical techniques. Em: **Machine learning, neural and statistical classification**. [s.l: s.n.]. p. 29–49.

MOLNAR, C. **Interpretable Machine Learning: A Guide for Making Black Box Models Explainable**. 2. ed. [s.l: s.n.].

MONTAGNER, C. C. et al. Ten years-snapshot of the occurrence of emerging contaminants in drinking, surface and ground waters and wastewaters from São Paulo State, Brazil. **Journal of the Brazilian Chemical Society**, v. 30, n. 3, p. 614–632, 2019.

- MONTAGNER, C. C.; VIDAL, C.; ACAYABA, R. Contaminantes emergentes em matrizes aquáticas do Brasil: cenário atual e aspectos analíticos, ecotoxicológicos e regulatórios. **Química Nova**, 11 jul. 2017.
- MORENO, R. A. et al. Anticonvulsivantes e antipsicóticos no tratamento do transtorno bipolar. **Brazilian Journal of Psychiatry**, v. 26, p. 37–43, 2004.
- MOURA, M. L. et al. Effect on Antimicrobial Resistance of a Policy Restricting Over-the-Counter Antimicrobial Sales in a Large Metropolitan Area, São Paulo, Brazil. **Emerging Infectious Diseases**, v. 28, n. 1, p. 180–187, jan. 2022.
- MURIUKI, C. et al. Mass loading, distribution, and removal of antibiotics and antiretroviral drugs in selected wastewater treatment plants in Kenya. **Science of The Total Environment**, v. 743, p. 140655, nov. 2020.
- NETSHITHOTOLE, R. et al. Occurrence of selected pharmaceuticals in wastewater and sludge samples from wastewater treatment plants in Eastern Cape province of South Africa. **South African Journal of Chemistry**, v. 78, p. 7–14, 2024.
- NEVES E CASTRO, P. B. et al. Antibiotic consumption in developing countries defies global commitments: an overview on Brazilian growth in consumption. **Environmental Science and Pollution Research**, v. 27, n. 17, p. 21013–21020, jun. 2020.
- NGUYEN, D.-V. et al. Assessing industrial wastewater effluent toxicity using boosting algorithms in machine learning: A case study on ecotoxicity prediction and control strategy development. **Environmental Pollution**, v. 341, p. 123017, jan. 2024.
- NHS. **NSAIDs. Non-steroidal anti-inflammatory drugs (NSAIDs) are medicines that are widely used to relieve pain, reduce inflammation, and bring down a high temperature.** Disponível em: <<https://www.nhs.uk/conditions/nsaids/>>.
- NRMMC. **Australian guidelines for water recycling: Augmentation of drinking water supplies.** EPHC/NRMMC/AHMC Canberra, , 2008.
- OLDENKAMP, R. et al. Regulatory Risk Assessment of Pharmaceuticals in the Environment: Current Practice and Future Priorities. **Environmental Toxicology and Chemistry**, v. 43, n. 3, p. 611–622, 2022.

OLIVEIRA, R. M. A.; SANT'ANNA, Â. M. O. **Modelos de aprendizado de máquina para detecção de anomalias no processo de manufatura de pneus**. A contribuição da engenharia de produção para desenvolvimento sustentável das organizações: Cadeias Circulares, sustentabilidade e tecnologias. **Anais...** Em: XLIII ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO. Fortaleza, Ceará, Brasil.: Universidade Federal da Bahia, 16 out. 2023. Disponível em: <<https://repositorio.ufba.br/handle/ri/38881>>. Acesso em: 19 set. 2024

OLIVEIRA, T. M. A. et al. Occurrence and Environmental Risk Assessment of Contaminants of Emerging Concern in Brazilian Surface Waters. **Environmental Toxicology and Chemistry**, v. 43, n. 10, p. 2199–2210, 1 out. 2024.

ONU. **Objetivos de Desenvolvimento Sustentável**. Disponível em: <<https://brasil.un.org/pt-br/sdgs>>. Acesso em: 19 set. 2024.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Wastewater treatment: Wastewater treatment is the connection rates to sewage treatment**. Disponível em: <<https://www.oecd.org/en/data/indicators/wastewater-treatment.html>>. Acesso em: 18 out. 2024.

PANI, A. et al. Macrolides and viral infections: focus on azithromycin in COVID-19 pathology. **International Journal of Antimicrobial Agents**, v. 56, n. 2, p. 106053, ago. 2020.

PARIDA, V. K. et al. Emerging contaminants in wastewater: A critical review on occurrence, existing legislations, risk assessment, and sustainable treatment alternatives. **Journal of Environmental Chemical Engineering**, v. 9, n. 5, p. 105966, 2021.

PATEL, M. et al. Pharmaceuticals of Emerging Concern in Aquatic Systems: Chemistry, Occurrence, Effects, and Removal Methods. **Chemical Reviews**, v. 119, n. 6, p. 3510–3673, 27 mar. 2019.

PEREIRA, A. et al. Selected Pharmaceuticals in Different Aquatic Compartments: Part II—Toxicity and Environmental Risk Assessment. **Molecules**, v. 25, n. 8, p. 1796, 14 abr. 2020.

PESCARA, I. C. **Ocorrência e remoção de contaminantes emergentes por tratamentos convencionais de água e esgoto**. Tese (Instituto de Química)—Campinas - SP: Universidade Estadual de Campinas (Unicamp), 2014.

PIGGOTT, T. et al. Global status of essential medicine selection: a systematic comparison of national essential medicine lists with recommendations by WHO. **BMJ Open**, v. 12, n. 2, p. e053349, fev. 2022.

PIVETTA, R. C. et al. Tracking the occurrence of psychotropic pharmaceuticals in Brazilian wastewater treatment plants and surface water, with assessment of environmental risks. **Science of The Total Environment**, v. 727, p. 138661, jul. 2020.

PORTAL DE DADOS ABERTOS. **Venda de Medicamentos Controlados e Antimicrobianos - Medicamentos Industrializados**. Disponível em: <<https://dados.gov.br/dados/conjuntos-dados/venda-de-medicamentos-controlados-e-antimicrobianos---medicamentos-industrializados>>. Acesso em: 13 out. 2024.

PUBCHEM. **PubChem. Explore Chemistry**. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/>>. Acesso em: 19 set. 2024.

REICHERT, G. et al. Emerging contaminants and antibiotic resistance in the different environmental matrices of Latin America. **Environmental Pollution**, v. 255, p. 113140, 2019.

REIS, E. O. et al. Occurrence, removal and seasonal variation of pharmaceuticals in Brazilian drinking water treatment plants. **Environmental Pollution**, v. 250, p. 773–781, jul. 2019.

RICHARDSON, S. D.; TERNES, T. A. Water Analysis: Emerging Contaminants and Current Issues. **Analytical Chemistry**, v. 86, n. 6, p. 2813–2848, 18 mar. 2014.

RODRIGUES, D. A. D. S. et al. Occurrence of antibiotics, hormones and PFAs in surface water from a Nile tilapia aquaculture facility in a Brazilian hydroelectric reservoir. **Chemosphere**, v. 352, p. 141444, mar. 2024.

RODRIGUES, F. et al. Pharmaceuticals in urban streams: A review of their detection and effects in the ecosystem. **Water Research**, v. 268, p. 122657, jan. 2025.

RODRÍGUEZ-RODRÍGUEZ, C. E. et al. Occurrence of pharmaceuticals in Latin America: case study on hazard assessment and prioritization in Costa Rica.

Environmental Monitoring and Assessment, v. 196, n. 8, p. 739, ago. 2024.

ROQUETTI, M. H. Proposta para derivação de critérios para contaminantes ambientais da agricultura: relatório final. 2010.

SANTOS, A. V. et al. Occurrence and risk assessment of pharmaceutically active compounds in water supply systems in Brazil. **Science of the total environment**, v. 746, p. 141011, 2020.

SILVA, R. F. DA. **Avaliação de risco ambiental de fármacos**. Tese (Doutorado em Sistemas de Gestão Sustentáveis)—Niterói, RJ: Universidade Federal Fluminense, 2019.

SILVA, M. O. et al. Perfil de utilização de medicamentos psicotrópicos dispensados por farmácias públicas durante a pandemia da COVID-19. **Research, Society and Development**, v. 11, n. 7, p. e45911730269, 31 maio 2022.

SMAALI, A. et al. Photocatalytic-persulfate-oxidation for diclofenac removal from aqueous solutions: modeling, optimization and biotoxicity test assessment.

Chemosphere, v. 266, p. 129158, 2021.

SNIP, L. J. P. et al. Modelling the occurrence, transport and fate of pharmaceuticals in wastewater systems. **Environmental Modelling & Software**, v. 62, p. 112–127, dez. 2014.

SOARES, A. F. S.; SOUZA, L. P. Contaminação das águas de abastecimento público por poluentes emergentes e o direito à saúde. **Revista de Direito Sanitário**, v. 20, n. 2, p. 100–133, 2020.

SUBEDI, B. et al. Mass loading and removal of pharmaceuticals and personal care products including psychoactives, antihypertensives, and antibiotics in two sewage treatment plants in southern India. **Chemosphere**, v. 167, p. 429–437, jan. 2017.

TAOUFIK, N. et al. Removal of emerging pharmaceutical pollutants: A systematic mapping study review. **Journal of Environmental Chemical Engineering**, v. 8, n. 5, p. 104251, out. 2020.

- TAOUFIK, N. et al. The state of art on the prediction of efficiency and modeling of the processes of pollutants removal based on machine learning. **Science of The Total Environment**, v. 807, p. 150554, 2022.
- TOHARUDIN, T. et al. Boosting Algorithm to Handle Unbalanced Classification of PM_{2.5} Concentration Levels by Observing Meteorological Parameters in Jakarta-Indonesia Using AdaBoost, XGBoost, CatBoost, and LightGBM. **IEEE Access**, v. 11, p. 35680–35696, 2023.
- TRINDADE, P. R. C. M. et al. Consumo doméstico de antibióticos entre 2014 e 2020 na cidade de Aracaju-SE. **Revista Multidisciplinar em Saúde**, v. 4, n. 1, 2023.
- U. S. FOOD AND DRUG ADMINISTRATION. Environmental Assessment of Human Drug and Biologics Applications: Guidance for Industry. 1998.
- UNESCO. **Intergovernmental Hydrological Programme**. Disponível em: <<https://www.unesco.org/en/ihp>>. Acesso em: 19 set. 2024.
- VAKARELSKA, E. et al. Persistent organic pollutants (POPs) - QSPR classification models by means of Machine learning strategies. **Chemosphere**, v. 287, n. Part 2, p. 132189, 2022.
- VAKILI, M. et al. Cross-linked chitosan/zeolite as a fixed-bed column for organic micropollutants removal from aqueous solution, optimization with RSM and artificial neural network. **Journal of environmental management**, v. 250, p. 109434, 2019.
- VALDEZ-CARRILLO, M. et al. Pharmaceuticals as emerging contaminants in the aquatic environment of Latin America: a review. **Environmental Science and Pollution Research**, v. 27, n. 36, p. 44863–44891, 2020.
- VERAS, T. B. et al. Analysis of the presence of anti-inflammatories drugs in surface water: A case study in Beberibe river - PE, Brazil. **Chemosphere**, v. 222, p. 961–969, maio 2019.
- VERGILI, I. et al. Occurrence and Prioritization of Pharmaceutical Active Compounds in Domestic/Municipal Wastewater Treatment Plants. **Bulletin of Environmental Contamination and Toxicology**, v. 102, n. 2, p. 252–258, fev. 2019.
- VERLICCHI, P. et al. Comparison of measured and predicted concentrations of selected pharmaceuticals in wastewater and surface water: A case study of a catchment

area in the Po Valley (Italy). **Science of The Total Environment**, v. 470–471, p. 844–854, 2014.

VERLICCHI, P.; AL AUKIDY, M.; ZAMBELLO, E. Occurrence of pharmaceutical compounds in urban wastewater: Removal, mass load and environmental risk after a secondary treatment—A review. **Science of The Total Environment**, v. 429, p. 123–155, jul. 2012.

VILLALOBOS, E. S. T. et al. Nueve años de tendencia en la resistencia a ciprofloxacina por *Escherichia coli*: estudio transversal en un hospital de Colombia. **Cadernos de Saúde Pública**, v. 40, p. e00031723, 2024.

WANG, D. et al. Towards better process management in wastewater treatment plants: Process analytics based on SHAP values for tree-based machine learning methods. **Journal of Environmental Management**, v. 301, p. 113941, jan. 2022.

WANG, Y.; FENNER, K.; HELBLING, D. E. Clustering micropollutants based on initial biotransformations for improved prediction of micropollutant removal during conventional activated sludge treatment. **Environmental Science: Water Research & Technology**, v. 6, n. 3, p. 554–565, 2020.

WILKINSON, J. L. et al. Pharmaceutical pollution of the world's rivers. **Proceedings of the National Academy of Sciences**, v. 119, n. 8, p. e2113947119, 2022.

WIRTZ, F. Danube, Meuse and Rhine MEMORANDUM 2008. **Environmental Science and Pollution Research**, v. 16, n. Suppl 1, p. 112–115, 2009.

WORLD HEALTH ORGANIZATION. World report on knowledge for better health: strengthening health systems. 2004.

WORLD HEALTH ORGANIZATION. Pharmaceuticals in drinking-water. p. 35, 2012.

WORLD HEALTH ORGANIZATION. Access to medicines: making market forces serve the poor. **World Health Organization**, 2017.

WORLD HEALTH ORGANIZATION. **Guidelines for drinking-water quality: incorporating the first and second addenda**. [s.l.] World Health Organization, 2022.

WU, C.; LI, B.; XIONG, N. An Effective Machine Learning Scheme to Analyze and Predict the Concentration of Persistent Pollutants in the Great Lakes. **IEEE Access**, v. 9, p. 52252–52265, 2021.

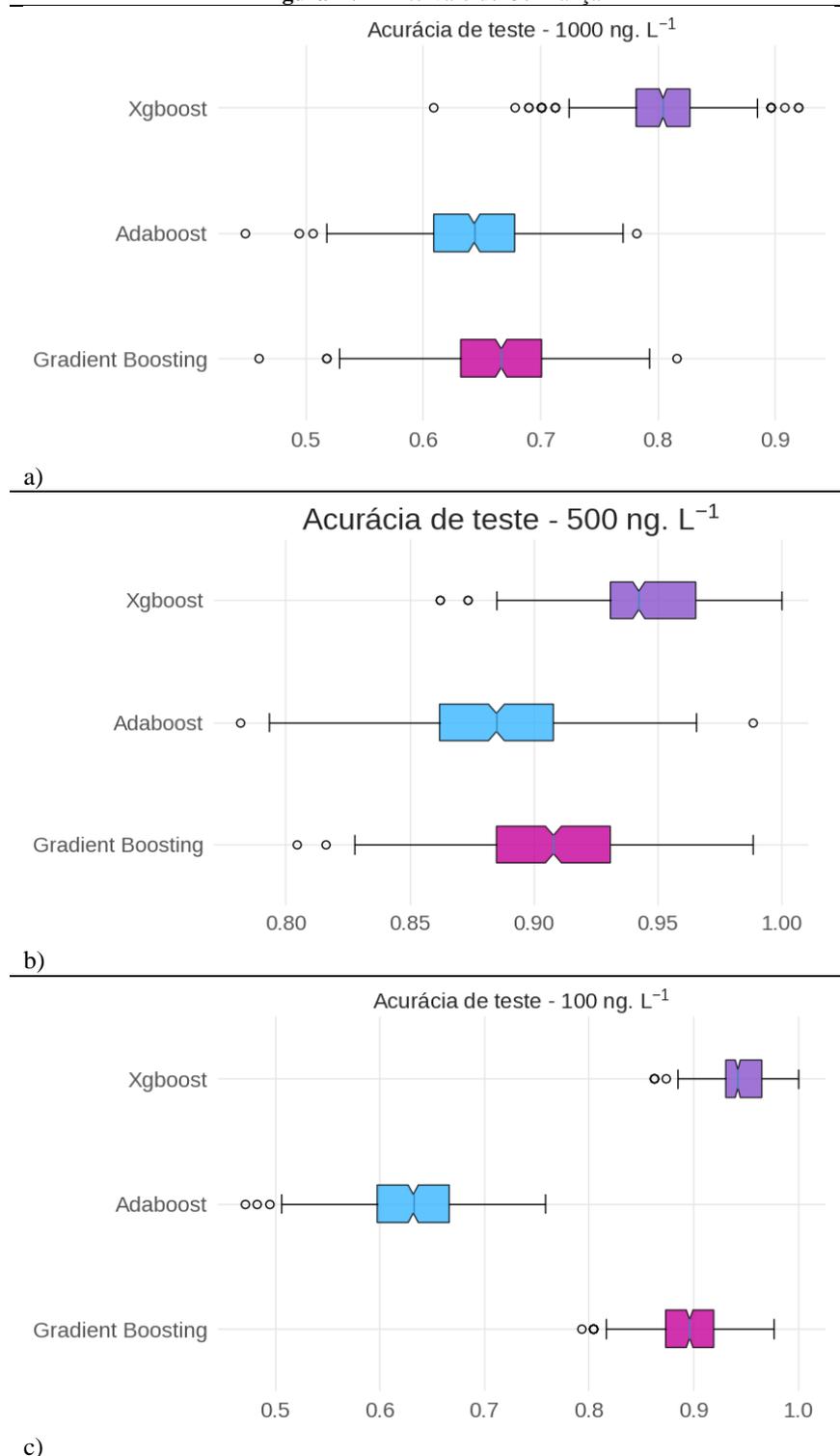
YAQUB, M. et al. Predictive modeling of pharmaceutical product removal by a managed aquifer recharge system: Comparison and optimization of models using ensemble learners. **Journal of Environmental Management**, v. 324, p. 116345, dez. 2022.

ZHU, X. et al. Machine learning for the selection of carbon-based materials for tetracycline and sulfamethoxazole adsorption. **Chemical Engineering Journal**, v. 406, p. 126782, fev. 2021.

APÊNDICE A – AVALIAÇÃO DOS MODELOS

MODELOS DE CARBAMAZEPINA

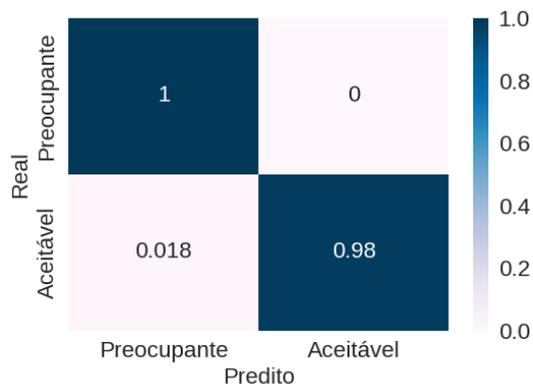
Figura A.1 - Intervalo de Confiança



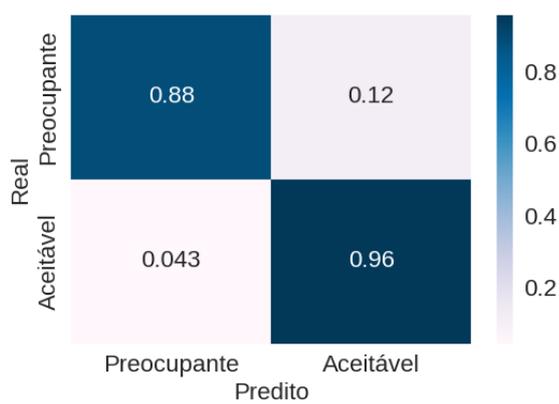
a) Ajustes de 1.000 ng. L⁻¹; b) Ajustes de 500 ng. L⁻¹; c) Ajustes de 100 ng. L⁻¹.

MODELOS DE ANTIBIÓTICOS

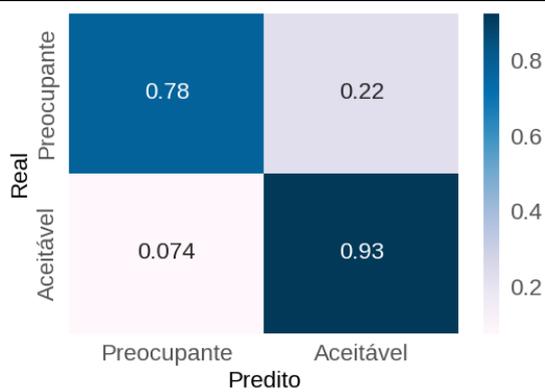
Figura A.2 – Matriz de Confusão



a)



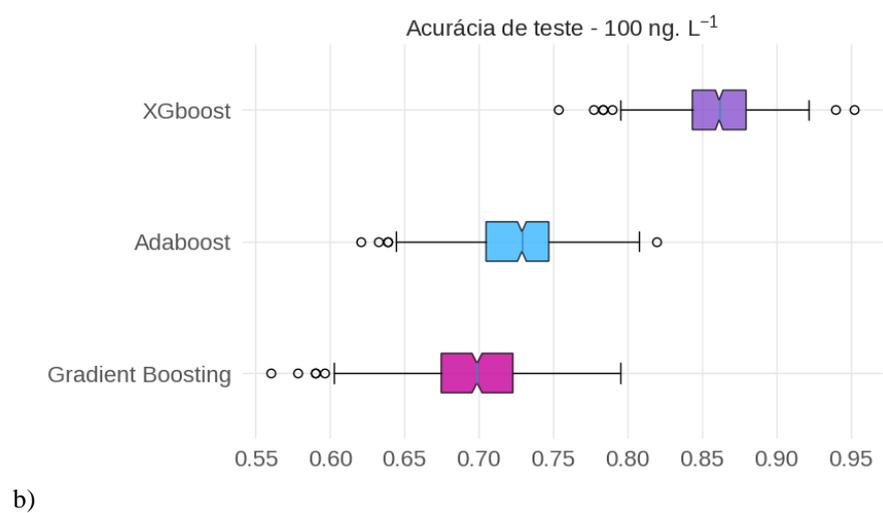
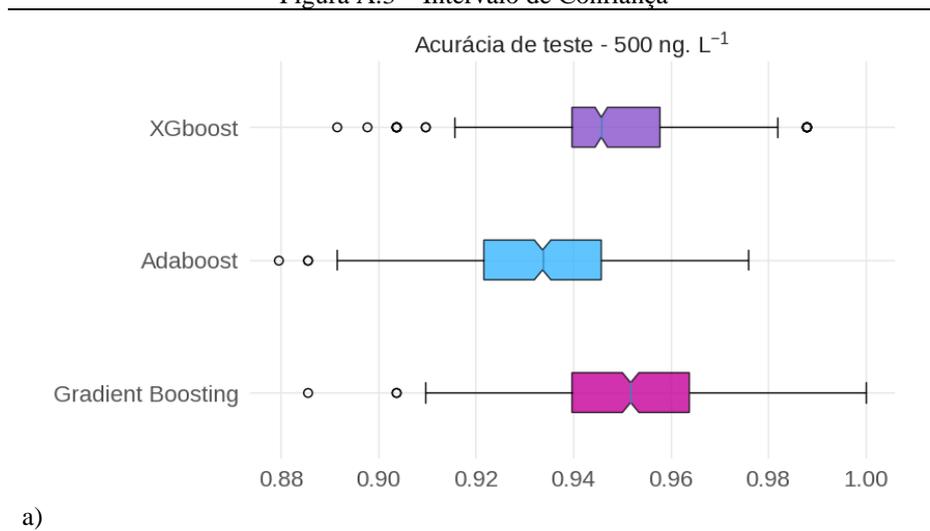
b)



c)

a) XGBoost, 1.000 ng. L⁻¹; b) XGBoost, 500 ng. L⁻¹; c) XGBoost, 100 ng. L⁻¹.

Figura A.3 – Intervalo de Confiança



a) XGBoost, 500 ng. L⁻¹; b) XGBoost, 100 ng. L⁻¹

ANEXO A – Dicionário de dados abertos da ANVISA

Campo	Descrição
ANO_VENDA	Ano da venda do medicamento.
MES_VENDA	Mês da venda do medicamento.
UF_VENDA	Unidade Federativa do endereço da farmácia ou drogaria, cadastrado no banco de dados da ANVISA, representando a UF onde ocorreu a venda.
MUNICIPIO_VENDA	Município do endereço da farmácia ou drogaria, cadastrado no banco de dados da ANVISA, representando o Município onde ocorreu a venda.
PRINCIPIO_ATIVO	Nome do princípio ativo do medicamento industrializado, conforme cadastrado no registro do medicamento, no banco de dados da ANVISA. Quando um medicamento tem mais de um princípio ativo, cada um deles é separado pelo caractere "+". Ex.: "PRINCIPIO ATIVO 1 + PRINCÍPIO ATIVO 2"
DESCRICA0_APRESENTACAO	Uma Apresentação de Medicamento representa o modo como um medicamento é apresentado na embalagem. Medicamento X, pode ter duas apresentações diferentes: • Apresentação 1: Uma caixa com 1 blister de alumínio com 20 comprimidos, cada comprimido com 5 mg de princípio ativo. Nesse caso, a descrição da apresentação seria: "5 MG COM CT BL AL X 20" • Apresentação 2: Uma caixa com 1 frasco de vidro com 50 mL de um xarope, com concentração do princípio ativo de 15 mg por mL. Nesse caso, a descrição da apresentação seria: 15MG/ML XPE CT FR VD x 50 ML Esses exemplos representam descrições de apresentações diferentes para um mesmo medicamento. Os termos utilizados na descrição das apresentações seguem o disposto no Vocabulário Controlado da ANVISA, disponível no link: http://portal.ANVISA.gov.br/documents/33836/2501339/Vocabu1%C3%A1rio+Controlado/fd8fdf08-45dc-402a-8dcffbb3fd21ca75
QTD_VENDIDA	Quantidade vendida de caixas ou frascos do medicamento.
UNIDADE_MEDIDA	Indica se a quantidade vendida do medicamento foi de caixas ou frascos.
CONSELHO_PRESCRITOR	Conselho de Classe do profissional que prescreveu o medicamento vendido.
UF_CONSELHO_PRESCRITOR	Unidade Federativa do Conselho de Classe do profissional que prescreveu o medicamento vendido.
TIPO_RECEITUARIO	Tipo de receituário utilizado na prescrição.
CID10	Classificação Internacional de Doença (aplicável apenas a medicamentos antimicrobianos).
SEXO	Sexo do paciente (aplicável apenas a medicamentos antimicrobianos).
IDADE	Valor numérico que representa a idade do paciente, em meses ou anos (aplicável apenas a medicamentos antimicrobianos).
UNIDADE_IDADE	Unidade de medida da idade do paciente, que pode ser em meses ou anos (aplicável apenas a medicamentos antimicrobianos).

Fonte: ANVISA (2020, p. 6).