



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE CIÊNCIAS DA SAÚDE  
DEPARTAMENTO DE CIÊNCIAS FARMACÊUTICAS  
CURSO DE GRADUAÇÃO EM FARMÁCIA

Vinicius dos Santos Machado

**Fármacos aprovados pela FDA e ANVISA desenvolvidos com o uso de  
Inteligência Artificial, uma revisão sistemática.**

Florianópolis  
2023

Vinicius dos Santos Machado

**Fármacos aprovados pela FDA e ANVISA desenvolvidos com o uso de  
Inteligência Artificial, uma revisão sistemática.**

Trabalho de Conclusão de Curso submetido ao curso de farmácia do Centro de Ciências da Saúde da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Farmacêutico.

Orientadora: Prof. Lilian Sibelle Campos Bernardes



Machado , Vinicius dos Santos

Fármacos aprovados pela FDA e ANVISA desenvolvidos com o uso de Inteligência Artificial, uma revisão sistemática. / Vinicius dos Santos Machado ; orientadora, Lilian Sibelle Campos Bernardes , 2024.

39 p.

Trabalho de Conclusão de Curso (graduação) -  
Universidade Federal de Santa Catarina, Centro de Ciências da Saúde, Graduação em Farmácia, Florianópolis, 2024.

Inclui referências.

1. Farmácia. 2. Farmácia . 3. Novos Fármacos . 4. Inteligência Artificial . 5. Agências Regulatórias . I. Bernardes , Lilian Sibelle Campos . II. Universidade Federal de Santa Catarina. Graduação em Farmácia. III. Título.



## RESUMO

A inteligência artificial (IA) vem revolucionando o desenvolvimento de fármacos e medicamentos, pela capacidade de acelerar e otimizar o processo como um todo, diminuindo custos e trazendo um novo medicamento ao mercado mais rapidamente. Esse estudo é uma revisão sistemática concentrada em medicamentos aprovados pela Food and Drug Administration (FDA) e Agência Nacional de Vigilância Sanitária (ANVISA) que utilizaram IA durante o desenvolvimento. A revisão empregou uma abordagem sistemática com análise bibliométrica, utilizando documentos regulatórios disponíveis nas bases de dados dessas agências. Foi desenvolvido um algoritmo utilizando a linguagem de programação Python para triagem de dados. O período de busca foi de 2015 a 2023, com um total de 398 medicamentos aprovados nesse período, tendo como critério de exclusão medicamentos em que não foi comprovado o uso de ferramentas de IA nos documentos analisados. Os objetivos específicos do estudo foram identificar as ferramentas de IA utilizadas no desenvolvimento de medicamentos e em que estágio elas foram aplicadas, bem como determinar quais dos medicamentos aprovados pela FDA que utilizavam IA também foram aprovados pela ANVISA. Ao final, o estudo resultou em 149 medicamentos triados pelo algoritmo com possíveis utilizações de IA, e após análise manual dos arquivos, dois foram comprovados, um deles sendo o Veklury® (remdesivir), aprovado tanto pela FDA quanto pela ANVISA, tendo utilizado *machine learning*, e o Brenzavvy® (bexagliflozina) aprovado somente pela FDA, utilizando *natural language processing*.

**Palavras-chave:** Inteligência artificial; FDA; ANVISA; *Machine Learning*; *Natural language processing*; Fármacos; Novos medicamentos.

## ABSTRACT

Artificial intelligence (AI) has been revolutionizing the development of drugs and medicines, through its ability to accelerate and optimize the process as a whole, reducing costs and bringing a new medicine to the market faster. This study is a systematic review of medicines approved by the Food and Drug Administration (FDA) and the Agência Nacional de Vigilância Sanitária (ANVISA) that used AI during development. The review employed a systematic approach with bibliometric analysis, using regulatory documents available in the databases of these agencies. An algorithm was developed using the Python programming language for data screening. The search period was from 2015 to 2023, with a total of 398 medicines approved during this period, with the exclusion criteria being medicines in which the use of AI tools was not proven in the documents analyzed. The specific objectives of the study were to identify the AI tools used in drug development and at what stage they were applied, as well as to determine which of the FDA-approved drugs that used AI were also approved by ANVISA. In the end, the study resulted in 149 drugs screened by the algorithm with possible uses of AI, and after manual analysis of the files, two were proven, one of them being Veklury® (remdesivir), approved by both the FDA and ANVISA, using a machine learning, and Brenzavvy® (bexagliflozin) only approved by the FDA, using natural language processing.

**Keywords:** Inteligência artificial; FDA; ANVISA; *Machine Learning*; *Natural language processing*; Fármacos; Novos medicamentos.

## LISTA DE FIGURAS

Figura 1. Etapas para obtenção dos arquivos.	5
Figura 2. Função utilizada para a pesquisa dentro dos arquivos	8
Figura 3. Estrutura química do remdesivir e seus metabólitos.	15
Figura 4. Aplicações do machine learning no desenvolvimento de fármacos.	20
Figura 5. Diagrama da inteligência artificial.	21
Figura 6. Estrutura de uma rede neural.	22
Figura 7. Estrutura da Bexagliflozina	24



## LISTA DE TABELAS

Tabela 1. Resultados da triagem dos arquivos.	10
Tabela 2. Resultados da triagem dos arquivos, contendo nomes dos medicamentos encontrados.	11

## SUMÁRIO

<b>1. INTRODUÇÃO</b>	1
1.1 OBJETIVO	4
1.1.1 GERAL	4
1.1.2 ESPECÍFICO	4
<b>2. METODOLOGIA</b>	4
<b>3. RESULTADO E DISCUSSÃO</b>	7
3.1 Levantamento dos Fármacos aprovados nas agências FDA e ANVISA	7
3.2 Análise dos Fármacos positivos	15
<b>4. CONCLUSÃO</b>	28
<b>REFERÊNCIAS</b>	28

## LISTA DE ABREVIATURAS E SIGLAS

ANVISA	Agência Nacional de Vigilância Sanitária
ATP	Trifosfato de Adenosina
CADD	<i>Computer-aided Drug Design</i> (Desenvolvimento de Fármacos Assistido por Computador)
CPU	<i>Central Processing Unit</i> (Unidade de Processamento Central)
CU	CASE Ultra
DL	<i>Deep Learning</i>
DM2	Diabetes Mellitus tipo 2
FDA	<i>Food and Drug Administration</i>
FOIA	<i>Freedom of Information Act</i> (Lei de Liberdade de Informação)
GPU	<i>Graphics processing Unit</i> (Unidade de processamento gráfico)
Hba1C	Hemoglobina Glicada
IA	Inteligência Artificial
LMA	<i>Leadscope Model Applier</i>
MCMC	Markov Chain Monte Carlo (Cadeia de Markov Monte Carlo)
ML	<i>Machine Learning</i>
OMS	Organização Mundial da Saúde
OS	<i>Operating System</i>
P&D	Pesquisa e Desenvolvimento
PCDT	Protocolo de diretrizes terapêuticas
PDB	<i>Protein Data Bank</i>
PNL	<i>Natural language Processing</i>
QSAR	<i>Quantitative structure-activity relationship</i> (Relação Estrutura-Atividade Quantitativa)
QSPR	<i>Quantitative Structure-Property Relationship</i> (Relação Estrutura-Propriedade Quantitativa)
REA	Relação Estrutura Atividade
RMN	Ressonância Magnética Nuclear
SGLT2	Cotransportador de Sódio-Glicose 2
SUS	Sistema Único de Saúde

## 1. INTRODUÇÃO

Ao longo dos anos, a Pesquisa e Desenvolvimento (P&D) de medicamentos tem sido considerado um processo longo, caro e com alto índice de falhas. Os principais desafios enfrentados pela indústria farmacêutica são o tempo médio de 12 a 15 anos e o custo variando de US \$161 milhões a US \$4,54 bilhões para desenvolver um medicamento (SCHLANDER, 2021). Além disso, a taxa de falha nos ensaios clínicos ultrapassa os 90% após os estudos pré-clínicos, e apenas 7% dos compostos passam por todas as etapas de ensaios clínicos e são aprovados pela agência de regulamentação de alimentos e medicamentos dos Estados Unidos (*Food and Drug Administration*, FDA) (DOWDEN, 2019).

O processo de P&D de um novo medicamento envolve encontrar compostos com atividade terapêutica, avaliar propriedades biológicas em testes *in vitro* e *in vivo* e realizar estudos clínicos em humanos. A primeira etapa do processo é a descoberta de compostos com potencial atividade terapêutica. Nessa fase, são realizadas pesquisas e experimentos em diversas fontes de compostos, para identificar moléculas com características desejáveis, como a capacidade de interagir com alvos moleculares específicos ou que demonstrem atividade celular associados à doença em questão, além da avaliação das propriedades físico-químicas e eventuais otimizações moleculares dos compostos promissores (FRIEDMAN, 2015).

Na segunda etapa, são realizados testes *in vitro* para avaliar as propriedades biológicas das moléculas selecionadas. Esses testes envolvem o uso de bioensaios, onde moléculas são testadas em um sistema biológico simulado ou modelo celular, para verificar sua atividade, eficácia e propriedades farmacocinéticas e farmacodinâmicas (FRIEDMAN, 2015).

A terceira etapa pode ser dividida em 4 fases: (i) estudos de fase I, os quais são conduzidos para estimar a tolerabilidade do fármaco, bem como para estudar propriedades farmacocinéticas e farmacodinâmicas em humanos; (ii) estudos de fase II, cujo objetivo principal é avaliar a eficácia e segurança do fármaco; (iii) estudos de fase III, nos quais a eficácia e segurança são avaliadas em um número maior de pacientes; (iv) estudos de fase IV que é a etapa de farmacovigilância (FRIEDMAN, 2015).

Os ensaios clínicos, quando bem delineados, com uma amostra suficiente de pacientes e países participantes, são considerados o padrão-ouro para determinar a eficácia e segurança de novos medicamentos. Graças aos ensaios clínicos, é possível obter dados confiáveis, que contribuem para a segurança da população em geral, reduzindo os efeitos colaterais e trazendo para o mercado um medicamento mais seguro e eficaz (FRIEDMAN, 2015).

Considerando a complexidade dos estudos clínicos ao longo de todas as suas fases, estima-se que os custos para a realização de estudos de fase I variam de aproximadamente US \$6,6 milhões a US \$1,4 milhões. Já para a fase II, os custos estimados são de US \$19,6 milhões a US \$7 milhões, e na fase III, os valores oscilam entre US \$52 milhões e US \$11,5 milhões (SERTKAYA, 2016).

A evolução dos métodos computacionais revolucionou a P&D, com o chamado *Computer-aided Drug Design* (CADD). Esse método permite identificar as interações entre os ligantes e respectivos alvos moleculares, rastreando um grande número de moléculas. Assim, selecionando apenas as mais promissoras para validação em laboratório, economizando em termos de custos e tempo.

Existem duas grandes vertentes do CADD: os métodos baseados em estruturas (*Structure-Based*) e os baseados nos ligantes (*Ligand-Based*). Os métodos baseados na estrutura dependem da estrutura química tridimensional do alvo e do ligante (o qual pode ser um composto ativo em estudo ou um fármaco já conhecido). Por outro lado, os métodos baseados nos ligantes requerem somente o conhecimento da estrutura química do ligante (SLIWOSKI, 2013).

A partir disso, novas ferramentas computacionais foram introduzidas no processo, como estudos de *docking* molecular, triagem virtual de alta eficiência, dinâmica molecular, uso de descritores moleculares, Relação Estrutura-Atividade Quantitativa (QSAR – do inglês *Quantitative Structure-Activity Relationship*), descoberta de fármacos baseada em fragmentos, *De Novo Drug Discovery*, dentre outras (SLIWOSKI, 2013).

No entanto, mesmo com moléculas muito promissoras, nem sempre elas apresentam as propriedades esperadas pelos órgãos regulatórios. Os principais motivos para o insucesso no desenvolvimento de novos medicamentos são: 79% segurança ou eficácia; 13% realinhamento estratégico, 7% razões comerciais; 1% incapacidade operacional ou técnica (DOWDEN, 2019).

Diante destes desafios, a busca por novos métodos que possam tornar o processo de P&D mais eficiente proporcionou o uso da inteligência artificial (IA) como uma abordagem promissora. Técnicas de *Machine Learning* (ML) e *Deep Learning* (DL), são usadas para analisar grandes volumes de dados e identificar padrões relevantes, gerando resultados valiosos em diferentes estágios do processo de desenvolvimento de medicamentos (RASHID, 2020).

A IA refere-se à inteligência demonstrada por máquinas, que imitam as capacidades inatas do ser humano, como criatividade, aprendizado, raciocínio e planejamento. É uma ciência abrangente, que inclui ciência da computação, cibernética, neurofisiologia, psicologia e linguística (LIANG, 2020).

Com o avanço da tecnologia de processamento de dados, algoritmos complexos e o crescimento exponencial de bibliotecas químicas, juntamente com o avanço na potência dos *hardwares*, a IA tem o potencial de desempenhar uma grande quantidade de tarefas em um curto espaço de tempo, podendo diminuir a duração total do desenvolvimento de um novo fármaco significativamente (DARA, 2021).

As ferramentas de IA podem ser utilizadas em várias etapas do desenvolvimento de fármacos, sendo elas: identificação de alvos farmacológicos, simulações e otimizações moleculares, predição de toxicidade e características físico-químicas, seleção de pacientes para ensaios clínicos; tratamento de dados, geração de compostos líderes, busca em diversas bases de dados; entre outros (LAMBERTI, 2019; CHUN, 2023).

Um exemplo é o fármaco DSP-1181, desenvolvido pela *Exscientia Pharmaceuticals* utilizando uma plataforma de IA, a qual é capaz de fazer um *screening* virtual em múltiplas bibliotecas de compostos e identificar quais são as mais promissoras. Toda a etapa de *screening* e pré-clínica foram feitas em menos de 12 meses, sendo que a média da indústria é entre 4 e 6 anos (BURKI, 2020).

A IA representa uma revolução na indústria farmacêutica, e conforme mais inovações são desenvolvidas nesta área, mais ela pode ser utilizada. Métodos, abordagens e estrutura das redes neurais são abundantes, e este trabalho busca identificar quais os fármacos aprovados pelo FDA e pela Agência Nacional de Vigilância Sanitária (ANVISA) foram desenvolvidos utilizando métodos de IA em alguma etapa de Pesquisa e Desenvolvimento.

## 1.1 OBJETIVOS

### 1.1.1 GERAL

Fazer um levantamento dos fármacos aprovados pelas agências FDA e ANVISA, cujo desenvolvimento empregou em alguma etapa ferramentas de inteligência artificial.

### 1.1.2 ESPECÍFICO

- Identificar quais fármacos aprovados pela FDA foram desenvolvidos utilizando ferramentas de inteligência artificial;
- Classificar os fármacos encontrados de acordo com a classe terapêutica;
- Identificar qual a ferramenta de IA utilizada e em qual etapa foi aplicada no desenvolvimento do fármaco;
- Verificar quais dos fármacos aprovados pela FDA que utilizaram IA foram aprovados pela ANVISA.

## 2 METODOLOGIA

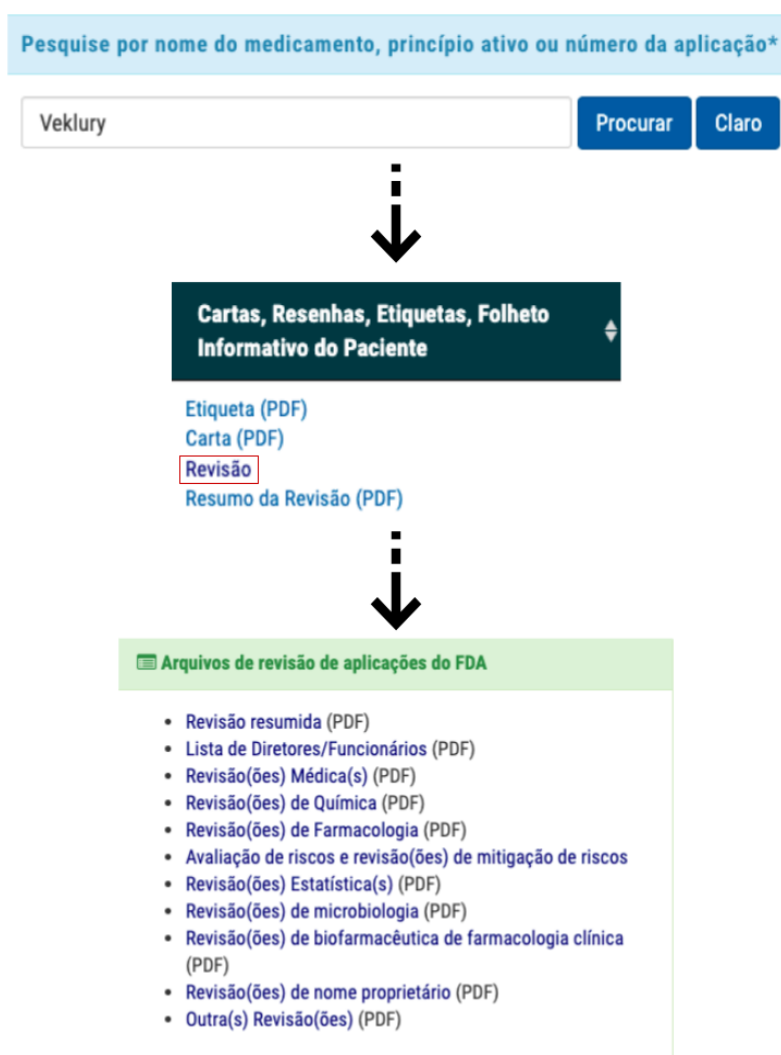
O presente trabalho trata de uma revisão sistematizada com análise bibliométrica, a qual foi realizada nos bancos de dados das Agências FDA e ANVISA. Para a triagem dos dados, foi criado um algoritmo utilizando a linguagem de programação Python. O período de busca foi de 2015 a 2023 e o critério de exclusão foram fármacos que não apresentaram nenhuma citação nos documentos analisados quanto ao uso de alguma ferramenta de IA.

Os dados foram obtidos por meio da busca do nome de cada medicamento na plataforma Drugs@FDA. Essa plataforma encontra-se disponível no próprio site da agência reguladora ou como um aplicativo para dispositivos móveis, compatível com IOS e Android. Para esse estudo, foi utilizada a plataforma online. O nome de

cada medicamento aprovado também está presente no próprio site do FDA, separado pelo ano de aprovação.

Após realizada a pesquisa, na aba “*Approval Date(s) and History, Letters, Labels, Reviews for NDA (new drug application)*” clica-se em *review*. Essa ação abrirá uma nova aba, a qual consiste em uma série de *links* para *download* dos arquivos em PDF referente a carta de aplicação final e no rótulo americano, na parte esquerda da página, e nas revisões das cartas de aplicação por parte da agência regulatória, na direita.

Figura 1 – Etapas para obtenção dos arquivos.



Fonte: Elaborado pelo autor, a partir de drug approval package: VEKLURY (2020).

Os arquivos utilizados foram: todas as revisões por parte do FDA das cartas de aplicação, que estão no quadro descrito como: "arquivos de revisão de



aplicações do FDA", como demonstrado na figura 1. Foi realizado o *download* dos arquivos que foram salvos em pastas separadas por ano, e pelos respectivos nomes dos medicamentos.

Posteriormente, a triagem dos dados foi realizada com o auxílio de um código na linguagem de programação Python. O *script* utiliza as bibliotecas, PyMuPDF e OS (*operating system*). Dentro da biblioteca PyMuPDF, foi utilizado o módulo FITZ.

O editor de código utilizado foi o Visual Studio code, nos sistemas operacionais macOS e Windows. No momento da escrita, também foi utilizado um copiloto, chamado github copilot. Esse programa está disponível para *download* diretamente no editor de código.

As palavras chave utilizadas nesse estudo foram: "*Machine learning*", "*Deep learning*", "*Artificial intelligence*", "*Computer vision*", "*Random Forest*", "*Natural language processing*", "*Bayesian*", "*TensorFlow*", "*Keras*", "*machine-learning*", "*deep-learning*", "*artificial-intelligence*", "*computer-vision*", "*random-forest*", "*natural-language-processing*", "*Neural Networks*", "*Supervised Learning*", "*Unsupervised Learning*", "*Semi-supervised Learning*", "*Transfer Learning*", "*Convolutional Neural Networks*", "*Natural Language Generation*", "*Chatbots*", "*Computer Speech Recognition*", "*Clustering*", "*Artificial Neural Network*", "*Big Data*", "*Data Mining*", "*AI Ethics*", "*big-data*", "*data-mining*", "*ai-ethics*", "*Convolutiuonal-neural-networks*", "*natural-lenguage-processing*", "*natural-language-generation*", "*computer-speech-recognition*", "*artificial-neural-network*", "*supervised-learning*", "*unsupervised-learning*", "*semi-supervised-learning*", "*transfer-learning*", "*neural-networks*", "*QSAR*", "*Quantitative Structure-Activity Relationship*", "*Qualitative Structure-Activity Relationship*", "*mcmc*", "*MCMC*", "*markov chain*", "*markov chain monte carlo*".

Após a execução do código, o *output* é dado em texto, conforme mostra-se a seguir: o arquivo X apresenta a palavra-chave X, para cada uma das respectivas palavras-chave encontradas. A partir disso, todos os arquivos desse *output* foram analisados em busca do contexto em que a palavra chave foi encontrada, utilizando o leitor de arquivos PDF, chamado Adobe Acrobat Reader.

Nesta etapa, novamente foi pesquisado a palavra-chave encontrada pelo *script*, utilizando a ferramenta interna de pesquisa do próprio programa. Se confirmado que o contexto condiz com a utilização de alguma ferramenta de

inteligência artificial, o medicamento é aceito. Assim, os medicamentos aceitos foram comparados à lista de aprovações de fármacos oferecida ao público pela ANVISA, disponível no site da própria agência.

Os dados encontrados estão apresentados em formato de tabelas e gráficos, além da discussão integrando os dados avaliados.

Por fim, para descrição, informação e discussão sobre as diferentes ferramentas de IA encontradas, foram utilizados artigos científicos disponibilizados nas bases de dados Web of Science, Pubmed e Google Scholar.

### **3 RESULTADOS E DISCUSSÃO**

#### **3.1 Levantamento dos Fármacos aprovados nas agências FDA e ANVISA**

Inicialmente foram identificados no site da FDA 398 fármacos aprovados no período em estudo e, a partir daí, para encontrar quais tiveram o uso de ferramentas de IA durante seu desenvolvimento, o fluxo de trabalho será apresentado em etapas para melhor entendimento do processo, como segue: coleta e organização dos documentos de cada fármaco; criação e utilização do código para identificação das palavras-chave utilizadas; análise manual dos dados obtidos pelo código.

##### **Coleta e Organização dos Documentos**

O nome de cada medicamento e o ano em que foi aprovado também foram obtidos no site da FDA. Os arquivos foram coletados na Plataforma Drugs@FDA, uma plataforma na qual é possível encontrar informações relacionadas ao processo de aprovação de novos fármacos. Através desta plataforma, a instituição responsável pelo desenvolvimento do medicamento, envia cartas de aplicação, as quais descrevem todo o processo de desenvolvimento e produção da nova entidade química, e recebem como resposta uma revisão sistemática por parte do FDA.

É importante ressaltar que especialistas em cada campo específico são os responsáveis por conduzir todas as fases da revisão, e se por acaso for identificada alguma irregularidade que afete diretamente a saúde do consumidor, o medicamento será reprovado e a empresa terá a obrigação legal de fazer as alterações

necessárias. Portanto, todo medicamento que é aprovado pelas agências regulatórias passaram por esse processo, o que garante a eficácia e segurança dos medicamentos comercializados no país.

Após o início da pesquisa, foi identificado que alguns medicamentos não apresentavam a totalidade dos arquivos para *download*. Isso causou uma limitação da pesquisa, pois a ausência desses arquivos pode resultar na exclusão involuntária do medicamento. Mesmo assim, foi obtido um total de 2874 arquivos em PDF.

Considerando que tais documentos são de propriedade privada da empresa detentora dos direitos do medicamento, via de regra o FDA não os disponibiliza ao público. Para obter acesso completo a todos os arquivos, é necessário solicitar ao governo americano por meio de uma requisição FOIA (*Freedom of Information Act*).

FOIA é uma lei publicada em 1966 e garante que qualquer pessoa ou organização possa ter acesso a dados confidenciais do governo americano, mesmo não sendo um cidadão ou organização americana (LEE, 2001).

### **Criação e utilização do código para identificação das palavras-chave utilizadas**

Antes de tudo, para que o código seja executado sem erros, duas etapas são necessárias, sendo elas: Definir a variável “pasta”, preenchendo a linha de código correspondente com o diretório da pasta principal onde os arquivos foram salvos primariamente. O segundo passo é o preenchimento do dicionário, com as palavras a serem pesquisadas. Com essas etapas concluídas, os arquivos são vasculhados página por página, utilizando a função apresentada na figura 2.

Figura 2 – Função utilizada para a pesquisa dentro dos arquivos.

```
for arquivo in os.listdir(pasta):
    if arquivo.endswith(".pdf"):
        caminho_arquivo = os.path.join(pasta, arquivo)
        with fitz.open(caminho_arquivo) as pdf:
            for page in pdf:
                conteudo = page.get_text().lower()
                for palavra_chave in palavras_chave:
                    if palavra_chave.lower() in conteudo:
                        arquivos_encontrados.append(arquivo)
                        break
```

Fonte: Elaborado pelo autor

A palavras chave, para serem encontradas, devem estar presentes no dicionário de palavras da maneira exata como está no arquivo a ser procurado, porém, sem levar em consideração as letras maiúsculas ou minúsculas. Por exemplo, se no código está presente a palavra chave “*machine learning*” e no texto estiver escrito “*machine-learning*”, o *script* não será capaz de identificar essa palavra.

Não existe limite quanto ao número de palavras-chave no dicionário, somente o tempo de processamento que aumenta conforme o aumento do número de palavras chave. Para fazer a leitura de todos os 2874 arquivos, que somados apresentam um tamanho de 14,0 GB, o tempo de processamento foi de 12 minutos.

Para o desenvolvimento de um código, é inato que se utilize bibliotecas de códigos, logo, as utilizadas nesse estudo foram as bibliotecas PyMuPDF e OS. Dentro de cada biblioteca, existem diferentes módulos, que podem ser importados individualmente, visando uma maior eficiência de *hardware* no momento de rodar o código.

Elas são essenciais para o desenvolvimento eficiente de *software*, pois permitem reutilizar códigos existentes e aproveitar o trabalho de outros desenvolvedores. A quantidade de bibliotecas disponíveis é imensa, e para cada tarefa, uma nova biblioteca pode ser importada.

A partir da biblioteca PyMuPDF (versão1.23.20), foi importado o módulo FITZ, que é utilizado quando se quer trabalhar especificamente com arquivos no formato de PDF. Com esse modulo é possível juntar arquivos, extrair, deletar, inserir, atualizar ou copiar informações presentes no texto. Sendo foi utilizado para a extração da respectiva palavra-chave encontrada no arquivo base.

Já a biblioteca OS (PYTHON SOFTWARE FOUNDATION), é importada quando é necessário utilizar funções que são dependentes do sistema operacional do próprio computador, sendo assim, foi utilizada para a navegação, abertura e leitura dos arquivos.

## OUTPUT

Após a execução do código, é gerado um *output*, que é a denominação utilizada para o resultado de um código. Dos 398 medicamentos aprovados, foram

então detectados 149 como possivelmente positivos para a utilização de inteligência artificial. Os resultados estão sumarizados nas tabelas 1 e 2.

Na tabela 1, os medicamentos foram separados por ano de aprovação e por palavra-chave encontrada. A coluna "Positivos", apresenta todos os fármacos daquele ano que foi identificado alguma palavra chave utilizada na busca. Nas colunas "QSAR" e "Bayes", estão medicamentos que apresentam essas palavras nos respectivos arquivos. Na coluna "Outros" estão todos os resultados de todas outras palavras chave.

Tabela 1 – Resultados da triagem dos arquivos.

Ano	Aprovados	Positivos	QSAR	Bayes	MCMC	Outros
2015	45	17 (37,8%)	5 (11,1%)	13 (28,9%)	2 (4,4%)	2 (4,4%)
2016	22	7 (31,8%)	2 (9,1%)	4 (18,2%)	2 (9%)	3 (13,6%)
2017	47	14 (29,8%)	4 (8,5%)	7 (14,9%)	3 (6,3%)	4 (8,5%)
2018	58	22 (37,9%)	9 (15,5%)	8 (13,8%)	2 (3,4%)	7 (12%)
2019	47	18 (38,3%)	7 (14,8%)	10 (21,2)	3 (6,3%)	4 (8,5%)
2020	53	18 (33,9%)	8 (15%)	10 (18,9%)	1 (1,8%)	4 (7,5%)
2021	47	26 (55,3%)	8 (13%)	12 (25,5%)	6 (12,7%)	7 (14,8%)
2022	43	9 (20,9%)	1 (2,3%)	7 (16,2%)	4 (9,3%)	4 (9,3%)
2023	36	18 (50%)	3 (8,3%)	14 (38,8%)	6 (16,6%)	7 (19,4%)
Total	398	149 (37,4%)	47 (11,8%)	85 (21,3%)	29 (7,2%)	42 (10,5%)

Fonte: elaborado pelo autor

Na tabela 2, os medicamentos foram novamente distribuídos quanto aos anos e palavras chave encontradas, sendo que nessa tabela, consta o nome dos medicamentos, facilitando assim a visualização dos resultados.

Tanto na tabela 1 quanto na 2, mais de uma palavra chave foi encontrada por medicamento, logo, o total de "Positivos" não condiz com a soma das outras colunas, melhor observado na tabela 2.

Tabela 2 – Resultados da triagem dos arquivos, contendo nomes dos medicamentos encontrados.

Ano	Positivos	QSAR	Bayes	MCMC	Outros
2015	Aristada® Avycaz® Cresemba® Daklinza® Empliciti® Farydak® Kybella® Nucala® Orkambi® Repatha® Rexulti® Savayasa® Strensiq® Upravi® Veltassa® Vraylar® Zurampic®  Total=17	Aristada® Daklinza® Kybella® Veltassa® Zurampic®  Total=5	Avycaz® Cresemba® Daklinza® Empliciti® Farydak® Nucala® Orkambi® Repatha® Savayasa® Strensiq® Upravi® Vraylar® Zurampic®  Total=13	Rexulti® Vraylar®  Total=2	
Ano	Positivos	QSAR	Bayes	MCMC	Outros
2016	Epclusa® Eucrisa® Ocaliva® Rubraca® Xiidra® Zepatier® Zinbryta®  Total=7	Ocaliva® Xiidra®  Total=2	Epclusa® Rubraca® Zepatier® Zinbryta®  Total=4	Eucrisa® Xiidra®  Total=2	Zepatier®  Total=1
Ano	Positivos	QSAR	Bayes	MCMC	Outros
2017	Besponsa® Bevyxxa® Brineura® Emflaza® Kevzara® Kisqali® Macrilen® Parsabiv® Radicava® Rhopressa® Vabomere® Xadago® Xepi® Xermelo®  Total=14	Bevyxxa® Macrilen® Xadago® Xermelo®  Total=4	Besponsa® Brineura® Kevzara® Kisqali® Parsabiv® Radicava® Vabomere®  Total=7	Kisqali® Rhopressa® Xepi®  Total=3	Emflaza®  Total=1

## Continuação da Tabela 2

Ano	Positivos	QSAR	Bayes	MCMC	Outros
2018	Aemcolo® Aimovig® Akynzeo® Annovera® Asparlas® Biktarvy® Braftovi® Emgality® Erleada® Mektovi® Moxidectin® Mulpleta® Olumiant® Orilissa® Pifeltro® Seysara® Symdeko® TPOXX® Xerava® Xofluza® Xospata®  Total=22	Aemcolo® Akynzeo® Biktarvy® Mektovi® Moxidectin® Pifeltro® Seysara® Symdeko® TPOXX®  Total=9	Asparlas® Braftovi® Emgality® Mektovi® Mulpleta® Orilissa® Xerava® Xospata®  Total=8	Aimovig® Olumiant®  Total=2	Aemcolo® Annovera® Orilissa® Symdeko® TPOXX®  Total=5
Ano	Positivos	QSAR	Bayes	MCMC	Outros
2019	Aklied® Brukinsa® Caplyta® Dayvigo® Feltroja® Inrebic® Jeuveau® Mayzent® Nourianz® Nubega® Polivy® Reblozyl® Reyvow® Rinvoq® Sunosi® Trikafta® Ubrely® Zulresso®  Total=18	Aklied® Caplyta® Reyvow® Rinvoq® Sunosi® Ubrely® Zulresso®  Total=7	Brukinsa® Dayvigo® Feltroja® Inrebic® Mayzent® Nourianz® Nubega® Polivy® Reblozyl® Trikafta®  Total=10	Aklied® Jeuveau® Sunosi®  Total=3	Nubeqa®  Total=1

Continuação da Tabela 2

Ano	Positivos	QSAR	Bayes	MCMC	Outros
2020	Bahrhemsys® Blenrep® Byfavo® Ebanga® Gavreto® Gemtesa® Inmazeb® Nexletol® Olinvyk® Orgovyx® Sogroya® Tabrecta® Tazverik® Tukysa® Veklury® Winlevi® Zeposia® Zokinvy®  Total=18	Byfavo® Olinvyk® Sogroya® Tazverik® Tukysa® Veklury® Zeposia® Zokinvy®  Total=8	Bahrhemsys® Blenrep® Byfavo® Ebanga® Gavreto® Inmazeb® Nexletol® Orgovyx® Sogroya® Tabrecta®  Total=10	Gemtesa®  Total=1	Veklury® Winlevi® Winlevi®  Total=3
Ano	Positivos	QSAR	Bayes	MCMC	Outros
2021	Adbry® Aduhelm® Azstarys® Brexafemme® Bylvay® Byfavo® Cabenuva® Evkeeza® Fotivda® Korsuva® Livtency® Lumakras® Livmarli® Leqvio® Lybalvi® Nexviazyme® Nulibry® Qelbree® Skytrofa® Tezspire® Truseltiq® Voxzogo® Nulibry®  Total=26	Brexafemme® Byfavo® Qelbree® Skytrofa® Livmarli® Cabenuva® Bylvay® Azstarys®  Total=8	Skytrofa® Truseltiq® Voxzogo® Nulibry® Lumakras® Leqvio® Korsuva® Fotivda® Evkeeza® Cabenuva® Byfavo® Aduhelm®  Total=12	Adbry® Korsuva® Livtency® Lybalvi® Nexviazyme® Tezspire®  Total=6	Nulibry®  Total=1



## Continuação da Tabela 2

Ano	Positivos	QSAR	Bayes	MCMC	Outros
2022	Cibinqo® Daxxify® Imjudo® Mounjaro® Omlonti® Quviviq® Relyvrio® Tzield® Sotyktu®  Total=9	Omlonti®  Total=1	Cibinqo® Imjudo® Mounjaro® Quviviq® Relyvrio® Tzield® Sotyktu®  Total=7	Cibinqo® Daxxify® Relyvrio® Tzield®  Total=4	
2023	Aphexda® Brenzavvy® Beyfortus® Dayblue® Epkinly® Filspari® Inpefa® Izervay® Jesduvroq® Litfulo® Leqembi® Miebo® Paxlovid® Qalsody® Rezzayo® Sohonos® Veozah® Xdemyvy®  Total=18	Inpefa® Sohonos® Veozah®  Total=3	Aphexda® Beyfortus® Brenzavvy® Dayblue® Epkinly® Filspari® Inpefa® Izervay® Sohonos® Jesduvroq® Leqembi® Litfulo® Rezzayo® Paxlovid®  Total=14	Brenzavvy® Dayblue® Filspari® Miebo® Qalsody® Xdemyvy®  Total=6	Brenzavvy®  Total=1
Total	149	47	85	26	13

Fonte: elaborado pelo autor

**Análise manual dos dados obtidos pelo código**

Após a obtenção dos resultados preliminares, foi observado que a maior prevalência nos resultados consiste em outras metodologias correlatas à IA, como estudos de QSAR, com 47 resultados (10,8%) e outros métodos estatístico-preditivos, como o bayesiano, com 85 resultados (21,3%) e a Cadeia de Markov Monte Carlo (MCMC, do inglês, *Markov Chain Monte Carlo*), com 26 (6,5%). Quanto às outras palavras chave encontradas, foi identificada a presença em somente 13 arquivos diferentes.

Esses métodos fundamentam a criação de alguns tipos de IA, como é o caso do *machine learning*. No entanto, não é possível afirmar que os métodos por si só constituem uma metodologia de IA, portanto, foram excluídos.

Já em relação aos modelos de IA, foram localizados em 2 arquivos diferentes, o primeiro, no medicamento Veklury®, que tem o princípio ativo chamado remdesivir, e o outro no Brenzavvy®, cujo princípio ativo é a bexagliflozina.

### 3.2 Análise dos Fármacos positivos

#### VEKLURY® - remdesivir

A palavra-chave detectada foi "*machine-learning*", no arquivo referente a revisão de farmacologia clínica, mais especificamente na seção onde discute-se sobre a toxicidade hepática e renal do próprio remdesivir e de mais dois de seus metabólitos, o trifosfato de remdesivir e o GS-441524, como apresentados na figura 3. (FDA, 2020)

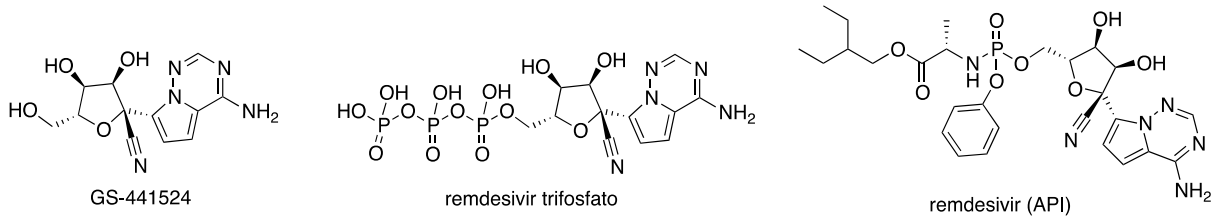
O arquivo começa detalhando a situação do processo de aprovação na data em que a carta foi escrita, pois como trata-se de um fármaco antiviral, cujo desenvolvimento primariamente estava focado na utilização contra o vírus ebola, foi constatada eficácia *in vitro* também contra o vírus SARS-CoV-2. O fármaco foi então aprovado em caráter emergencial, no dia 1 de maio de 2020, por parte do FDA e em seguida por várias outras agências regulatórias pelo mundo. (LAMB, 2020)

Causador da doença Covid-19, também conhecido como corona vírus, o SARS-CoV-2, em 30 de janeiro de 2020, causou oficialmente uma pandemia, segundo a organização mundial da saúde (OMS), deixando o mundo inteiro em busca da cura, pois todas as alternativas utilizadas até o momento mostraram-se insuficientes. (GUO, 2019).

Em meio ao cenário de pandemia, o remdesivir surgiu como uma alternativa para o tratamento de pacientes internados em estado grave. Porém, já era conhecido que o fármaco poderia causar falha hepática, por conta de vários casos reportados, em que as atividades das transaminases apresentaram-se aumentadas. No entanto, um dos desfechos clínicos da própria Covid-19 são lesões hepáticas,

deixando inconclusiva a hipótese relacionada a ação adversa do fármaco e se os casos reportados realmente apresentavam algum significado clínico. (LAMB, 2020)

Figura 3 - Estrutura química do remdesivir e seus metabólitos.



Fonte: Adaptado de Veklury. (2020)

Como o remdesivir se trata de um pró-fármaco, ou seja, a molécula não é ativa farmacologicamente na estrutura molecular original, necessitando da metabolização após administração, logo não há permanência na corrente sanguínea tempo suficiente para causar efeitos tóxicos. Já os metabólitos ativos do fármaco sim, apresentando potencial de causar danos na vesícula biliar. (FDA, 2020)

O mecanismo farmacológico do remdesivir dá-se pela inibição da síntese de RNA viral, por meio da incorporação de um metabólito ativo, o trifosfato de remdesivir, no sítio de ligação do trifosfato de adenosina (ATP) da RNA polimerase viral. Nesse sítio da enzima o metabólito atua como um análogo competitivo de alta seletividade. Logo, quando o fármaco está biodisponível, liga-se 3,65 vezes mais do que o próprio ATP e nestas condições, a RNA polimerase fica inativada. (LEVIEN, 2021)

Esses resultados foram obtidos por meio de duas plataformas estatísticas chamadas *Leadscope Model Applier* (LMA) e *CASE Ultra* (CU). Ambas são plataformas preditivas baseadas em estatística para a obtenção de estudos de relação estrutura-atividade quantitativos (QSAR, do inglês *Quantitative structure-activity relationship*) desenvolvidos por meio de *machine learning* (FDA, 2020).

Estudos *in vitro* também foram realizados para avaliar outros métodos de toxicidade. O resultado foi positivo para interação com a DNA polimerase, mas sendo esse um mecanismo que necessita um tempo grande de exposição e uso constante. O risco foi avaliado como clinicamente não relevante (FDA, 2020).

Sendo assim, o uso do remdesivir foi aprovado pela FDA, porém com a ressalva de que a saúde hepática do paciente deveria ser monitorada antes do início do tratamento e durante a administração e internação do paciente (FDA, 2020).

Em relação a ANVISA, o Veklury® foi aprovado em março de 2021, sendo indicado inicialmente para pacientes adultos e adolescentes com pneumonia causado por corona vírus que requerem a administração de oxigênio. Um ano depois, em maio de 2022, o uso do medicamento foi ampliado para pacientes adultos e pediátricos que necessitem ou não da administração suplementar de oxigênio, ampliando assim o escopo de pacientes. (Veklury, 2022)

A utilização da IA nesse cenário foi de grande utilidade, visto que, em decorrência da situação de pandemia incluindo o distanciamento social, experimentos em laboratório não poderiam ser conduzidos.

#### **QSAR (RELAÇÃO ESTRUTURA-ATIVIDADE QUANTITATIVA, DO INGLÊS, QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP)**

Estudos de QSAR é a investigação das conexões entre fragmentos da estrutura molecular de fármacos e a respectiva participação na atividade, inatividade, toxicidade, absorção, seletividade, etc. Baseando-se em vários princípios fundamentais como: similaridade, linearidade e aditividade. Com o QSAR, é possível prever de maneira numérica as atividades biológicas de moléculas, pela obtenção de gráficos de tendência e comparações entre os valores dos descritores moleculares em relação às atividades biológicas. (TROPSHA, 2010)

Os modelos QSAR são tipicamente modelos de regressão ou classificação, que relacionam um conjunto de variáveis preditoras a uma variável de resposta. Essas variáveis são tipicamente as propriedades intrínsecas das moléculas, que vão desde a estrutura como, propriedades eletrônicas, representação de grupos funcionais ou características geométricas, até propriedades físico-químicas, como peso molecular, área de superfície polar, coeficiente de partição ou até a participação na atividade ou inatividade farmacológica, a toxicidade ou absorção dentre os mais variados tecidos ou células, etc. (TROPSHA, 2010)

Um conceito semelhante é a base das relações quantitativas estrutura-propriedade (QSPR). Geralmente, esses métodos QSAR ou QSPR tentam encontrar relações causais entre variáveis preditoras para um *endpoint* específico.

Porém, para que isso seja capaz, bases de dados confiáveis e tratadas são necessárias, como por exemplo o chEMBL, com 2,4 milhões de estruturas e o *Protein Data Bank* (PDB), com mais de 1 milhão. Nas bases de dados, é possível encontrar além da estrutura química, de forma detalhada e em 3D, os estudos sobre a atividade biológica de moléculas.

Como os fármacos devem ser preferencialmente moléculas pequenas, conforme indicado pelas 5 regras de Lipinski, “para uma molécula ser considerada um potencial fármaco, a estrutura deve apresentar o maior número possível entre as 5 regras, que são: o valor do log P superior ou igual a 5, a massa molecular inferior ou igual a 500, o número de aceitadores de ligação de hidrogênio menor ou igual a 10, e o número de doadores de ligação de hidrogênio menor ou igual a 5”. (BARREIRO, 2014)

Logo, a estrutura química de potenciais fármacos nem sempre é extremamente complexa, o que possibilita que os algoritmos tenham uma capacidade de previsão extremamente alta. Isso se deve ao fato de que, fármacos com estruturas similares facilitam a identificação de padrões estruturais relevantes, que poderiam ser ignorados caso fossem únicos ou apresentassem um formato não linear. (SERAFIM, 2021)

Esse fenômeno ocorre devido à disponibilidade de uma ampla variedade de estruturas, permitindo a seleção de um grande número de moléculas com a mesma atividade biológica no grupo de treinamento. Com essa abordagem, torna-se possível analisar diferentes partes da molécula e o respectivo impacto em uma característica específica, por meio de uma amostra significativa, o que aumenta a confiabilidade dos resultados. (SERAFIM, 2021)

Embora pareça simples, essa tarefa não é trivial. Durante as análises, vários problemas podem surgir, como a falta de qualidade dos dados. A capacidade de previsão dos modelos é diretamente baseada na qualidade dos dados fornecidos, e se entre os dados contiverem estruturas falsas, duplicadas ou com erros na disposição dos átomos e não venham a ser identificados, os resultados não serão confiáveis.

É por essas razões que os modelos de *machine learning* criados para uso médico e aprovados pelo FDA precisam disponibilizar a base de dados usada para treiná-los, a fim de detectar falhas e comprovar a confiabilidade do software. (FDA, 2023)

Um outro problema relevante na área química, são os resultados com moléculas que não podem ser sintetizadas. A síntese de fármacos, seja em escala industrial ou em bancada, é um processo extremamente complexo e, em alguns casos, até impossível. Como os algoritmos não são perfeitos, podem concentrar-se apenas em uma característica específica e negligenciam a rota sintética para a obtenção de uma molécula específica, criando grupos funcionais inexistentes, heterocíclicos incompatíveis ou estruturas impossíveis. (RASHID, 2021)

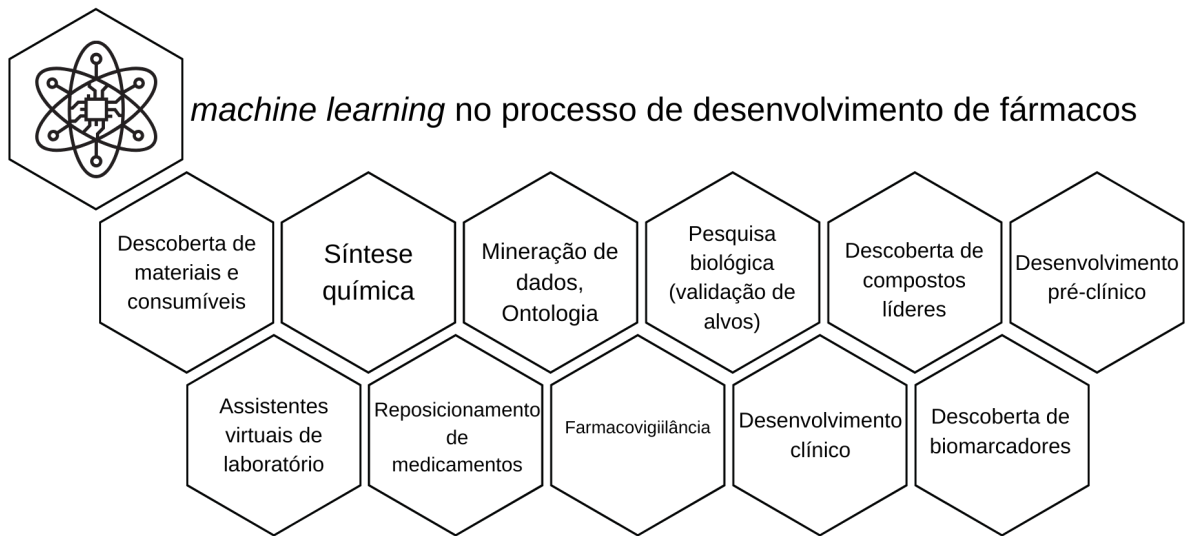
## **MACHINE LEARNING**

Métodos de *machine learning* são frequentemente utilizados para melhorar a precisão dos modelos de QSAR, pois basicamente, as diferentes abordagens utilizam como base o princípio de similaridade entre moléculas conhecidas, para extrapolar os dados a novas moléculas desconhecidas. Logo, quanto mais dados de qualidade e mais rodadas de treino, maior a confiabilidade e robustez do método. (DARA, 2021)

Não somente aplicados a estudos de QSAR, o *machine learning* pode ser utilizada em muitas outras etapas para acelerar o desenvolvimento de fármacos, como o *screening* de alta eficiência, obtenção de dados, otimização e obtenção de rotas sintéticas, reposicionamento de fármacos, farmacovigilância, validação de alvos farmacológicos, predição de dados, entre outros que se encontram sumarizados na figura 5. (DARA,2021)

Resumindo, em essência *machine learning* é um campo de estudos que envolve o desenvolvimento de algoritmos que permitem que computadores aprendam e façam decisões e previsões sem serem programados previamente para essa determinada função explicitamente, obtendo assim a capacidade de analisar e aprender a partir de grandes quantidades de dados. (DARA,2021)

Figura 4 – Aplicações do *machine learning* no processo de desenvolvimento de fármacos.



Fonte: Adaptado de DARA et al, 2021

Logo, técnicas utilizadas na neurociência também são utilizadas no campo da IA, como por exemplo aprendizado por reforço. Métodos de *reinforcement learning* utilizam várias rodadas de treino e conforme os resultados são obtidos, uma pontuação específica é dada para cada um deles. Essa abordagem leva a máquina a distinguir entre o certo e o errado, por meio dos resultados que a própria apresenta. (DARA,2021)

Essa técnica envolve tentativa e erro, incluindo o *feedback* de recompensas ou penalidades. Várias abordagens podem ser tomadas, como por exemplo: métodos baseados em valores, métodos baseados em políticas e métodos baseados em modelos. (DARA,2021)

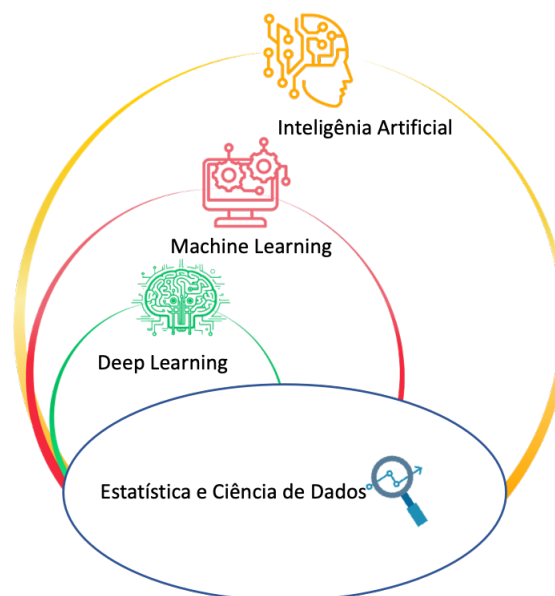
## DEEP LEARNING

Algoritmos de *deep learning* consistem em uma subdivisão do *machine learning* e estão entre os avançados, com resultados que ultrapassam a capacidade

de outros métodos. Entretanto, a estrutura e necessidade de *hardware* também são elevados. Técnicas convencionais de *machine learning* utilizam o processador, o chamado CPU (*central processing unity*), para realizar os cálculos necessários. Já no aprendizado profundo, é utilizada a placa gráfica, a chamada GPU (*graphics processing unit*). A maior diferença entre as duas reside na arquitetura e no foco de processamento. (ZHOU, 2019)

As Unidades Centrais de Processamento (CPUs, do inglês Central Processing Unit) têm o objetivo de executar tarefas gerais de maneira eficiente, utilizando núcleos de processamento de alto desempenho. Em contraste, as Unidades de Processamento Gráfico (GPUs, do inglês Graphics Processing Unit) foram desenvolvidas para maximizar a execução paralela, contando com milhares de núcleos de processamento menores. Essa característica faz com que as GPUs sejam ideais para lidar com tarefas que envolvem operações matriciais e cálculos em grande escala, cuja ocorrência é comum em algoritmos de aprendizagem de máquina. (DARA, 2021)

Figura 5 - Diagrama da inteligência artificial.



Fonte: Adaptado de ANDRADE, 2023.

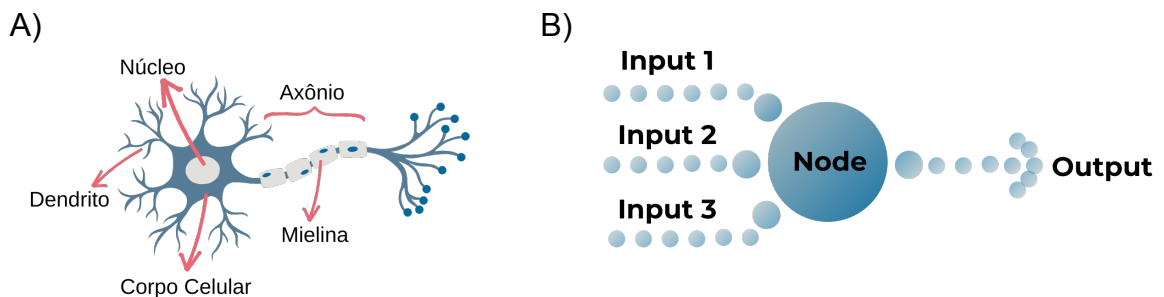
Uma outra distinção importante está relacionada à flexibilidade e programabilidade. As CPUs são altamente flexíveis e capazes de executar uma



ampla variedade de tarefas, porém geralmente com um desempenho inferior em cálculos massivos de aprendizado de máquina. Já as unidades de processamento gráfico (GPUs), por sua vez, possuem uma alta capacidade de programação, permitindo a execução de operações personalizadas de baixo nível, o que as torna ideais para algoritmos de aprendizado profundo, porém podem ser menos versáteis em tarefas de propósito geral. (DARA,2021)

No que diz respeito à programação, as GPUs oferecem a capacidade de definir e executar operações específicas, conhecidas como "*kernels*". Esses *kernels* são compostos por um conjunto de instruções que são executadas em paralelo por múltiplos núcleos da GPU, possibilitando o processamento eficiente de grandes conjuntos de dados, como os utilizados em cálculos de aprendizado de máquina. (DARA,2021)

Figura 6 – Estrutura de uma rede neural.



Fonte: Adaptado de MUTHUKRISHNAN (2020)

Algoritmos de *deep learning* utilizam-se de redes neurais, que podem ser definidas como: camadas de tomadas de decisão sequências que imitam o cérebro humano. São utilizadas, porém, com uma grande diferença em relação aos outros algoritmos, algumas dessas camadas apresentam camadas de dados invisíveis, as chamadas de *black box*, onde não é possível saber o que é gerado nessa etapa. (DARA, 2021)

Essas redes neurais funcionam de maneira interconectada, sendo que cada "neurônio" ou unidade recebe um *input* de entrada, realiza uma operação matemática a partir desse sinal e produz um *output* de saída. Assim, uma rede neural interconectada "aprende" com os *inputs* ajustando os valores dos

denominadores dessas operações matemáticas por meio de um treinamento. Esse treinamento envolve fornecer dados para a rede neural e comparar os resultados obtidos com os resultados esperados. (MUTHUKRISHNAN, 2020)

Um exemplo bastante relevante, que tem ganhado cada vez mais popularidade recentemente, é o Alpha Fold. Trata-se de um algoritmo desenvolvido pela empresa *deep mind*, que utiliza um método de *deep learning*. Esse algoritmo possui a incrível habilidade de prever a estrutura tridimensional de uma proteína, apenas analisando a sequência primária de aminoácidos. Diversos outros métodos de inteligência artificial, como *machine learning*, redes neurais convencionais e aprendizado por reforço, foram utilizados para solucionar essa mesma problemática. (NUSSINOV, 2022)

No entanto, nenhum desses métodos obteve resultados tão impressionantes quanto o *deep learning*, que conseguiu apresentar estruturas com confiabilidade superior a 90% em alguns casos. (NUSSINOV, 2022)

Esta ferramenta está solucionando um grande dilema ao possibilitar a obtenção de estruturas 3D confiáveis, sem a necessidade de técnicas mais caras e demoradas como a cristalografia de raios X, que ainda é considerada o padrão ouro. Atualmente, essa ferramenta está colaborando para o crescimento da base de dados do PDB, uma vez que, aproximadamente 80% das moléculas presentes utilizam essa abordagem preditiva, totalizando cerca de 800.000 unidades.

### **BRENZAVVY® - bexagliflozina**

A palavra-chave encontrada foi "*Natural Language Processing*", no arquivo de revisão integrativa. O arquivo contém 445 páginas e discute sobre várias etapas do desenvolvimento do fármaco, a IA foi utilizada para a pesquisa em bases de dados.

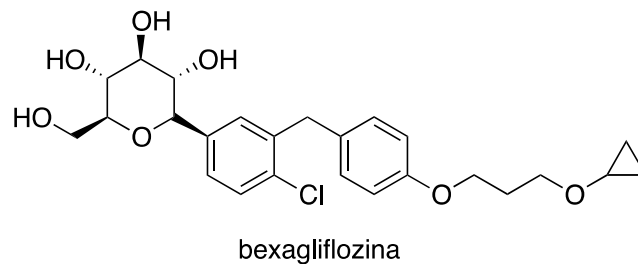
*Natural Language Processing (PNL)* é um ramo onde ocorre a intersecção da IA com a linguística, por meio de algoritmos capazes de interpretar, racionalizar e gerar linguagem humana. (NADKARNI, 2011)

O processamento de linguagem natural (PLN) foi utilizado no programa clínico da bexagliflozina para consultar dados de segurança em textos de livre acesso e identificar eventos de queda. (FDA, 2023)

Eventos de queda são descritos como eventos adversos associados à perda de controle postural, mas não fornecem detalhes sobre atividades específicas de queda. A fonte menciona que as quedas podem ter etiologias diversas, como hipoglicemia, hipotensão, convulsões, síncope, trauma ou eventos vestibulares. (FDA, 2023)

O Requerente utilizou abordagens baseadas em PNL para analisar narrativas de assuntos, termos literais e o banco de dados de segurança Argus, que rastreia e analisa eventos de segurança individuais. Esta abordagem permitiu a identificação de um número significativo de eventos de queda que podem ter sido subnotificados nos dados de segurança. (FDA, 2023)

Figura 7– Estrutura química da bexagliflozina



Fonte: Adaptado de DALAL et al, 2023.

Estes casos são importantes quando se fala de um medicamento utilizado para o tratamento de pacientes com diabetes mellitus do tipo 2 (DM2), doença onde em que o pâncreas não excreta insulina em quantidades suficientes para regular a taxa de glicose (HOY, 2023), e que trata-se também de um fármaco que age sobre a reabsorção renal de glicose, sendo assim, casos de hipoglicemia podem ser mais frequentes. (FDA, 2023)

O mecanismo de ação da bexagliflozina é por meio da inibição do cotransportador de sódio-glicose 2 (SGLT2), impedindo a reabsorção de glicose no túbulo proximal, aumentando assim a excreção de maiores taxas de glicose pela urina. (DALAL, 2023)

Segundo o protocolo de diretrizes terapêuticas (PCDT) para o tratamento da DM2, é indicado inicialmente a modificação de hábitos de vida juntamente com o fármaco denominado metformina, sendo essa a primeira escolha terapêutica. (MINISTÉRIO DA SAÚDE, 2020)

Estudos demonstram que a bexagliflozina juntamente com a metformina apresentam resultados significativos quanto a diminuição da hemoglobina glicada (HbA1c) em relação a pacientes que utilizam somente a metformina. (HALVORSEN, 2023)

O inibidor de SGLT2 disponível atualmente no sistema único de saúde (SUS) é a dapagliflozina, sendo recomendado para pacientes com DM2, idade igual ou superior a 65 anos e doença cardiovascular estabelecida, que não conseguiram fazer o tratamento com sulfonilureia juntamente com metformina. (MS, 2020)

A bexagliflozina apresenta uma revolução quanto ao tratamento da DM2, pois foi aprovado como monoterapia por parte do FDA, somente sendo orientado ao paciente a mudança de estilo de vida, envolvendo atividades físicas e dieta, não necessitando uso concomitante da metformina, tornando-se assim um possível novo fármaco de primeira escolha. (HOY, 2023)

Tendo sido aprovado somente em janeiro de 2023, mais estudos quanto à eficácia e seus benefícios ainda precisam ser realizados, explicando assim o porquê de sua aprovação perante a ANVISA ainda não ter sido feita.

## **NATURAL LANGUAGE PROCESSING**

O Processamento de Linguagem Natural (PLN) é um campo de estudo que concentra-se na interação entre computadores e a linguagem humana, permitindo que os computadores entendam, interpretem e gerem a linguagem humana. (NADKARNI,2011)

Tendo início na década de 1950, a PNL envolve o desenvolvimento de algoritmos e modelos que permitem aos computadores processar e analisar dados de texto, extrair significado e realizar tarefas como tradução de idiomas, análise de sentimentos, classificação de texto e extração de informações. A PNL abrange várias abordagens para análise de texto, incluindo reconhecimento de fala, processamento sintático e abordagens semânticas para recursos linguísticos. (NÉMETH,2023)

Ele desempenha um papel crucial em tarefas como resumo automatizado de texto, *chatbots*, assistentes de voz e compreensão de idiomas nos mecanismos de pesquisa. Um exemplo bastante presente no dia a dia é a Siri, assistente de voz presente no *Iphone*.

Os algoritmos de PNL podem detectar estruturas semânticas, conotações e significados latentes em grandes corpora, que vão além do conteúdo explícito e requerem um nível de compreensão e interpretação semelhante à inteligência humana. A PNL envolve o uso de métodos e algoritmos computacionais para processar e analisar dados textuais, que é um aspecto fundamental da pesquisa em IA. (MAHARASHTRA, 2023)

## 4 CONCLUSÃO

Ao fim deste trabalho, é possível observar que dentre os 398 medicamentos aprovados nos 9 anos pesquisados, somente 2 utilizaram IA, que foram o Veklury® (remdesivir), fármaco antiviral utilizado para o tratamento de pacientes com COVID-19, e o Brenzavvy® (bexagliflozina), inibidor do cotransportador SGLT2, aprovado como monoterapia para pacientes com diabetes tipo 2.

As técnicas foram respectivamente *machine learning*, utilizada no contexto de otimização de resultados de estudos de QSAR em duas plataformas estatístico-preditivas. E *natural language processing (PNL)* utilizada para a pesquisa no banco de dados de efeitos adversos, à procura de eventos de queda.

Ambas as utilizações foram importantes para o processo de desenvolvimento dos medicamentos, sendo que no caso do Veklury®, a aprovação foi concedida em caráter emergencial, necessitando uma agilidade no processo de obtenção de resultados. No caso do Brenzavvy®, havia a suspeita de subnotificação quanto aos eventos de queda por parte da requerente, logo, a técnica agilizou a pesquisa no banco de dados e aumentou a confiabilidade dos resultados.

É importante ressaltar que a IA ainda é mais utilizada para as etapas iniciais do processo de P&D, além do que, as técnicas ainda são bastante recentes, logo, os fármacos que as estão utilizando nos respectivos processos de desenvolvimento, ainda não foram aprovados, e se já foram, os arquivos ainda não estão disponíveis ao público.

Em relação à ANVISA, nenhum documento quanto ao desenvolvimento dos medicamentos é disponibilizado ao público em geral, dificultando a realização de pesquisas como essa utilizando bases de dados nacionais e em português. As informações disponibilizadas permitem identificar somente se o medicamento foi aprovado ou não. Até o momento somente o Veklury® foi aprovado por parte da ANVISA.

No que se refere à utilização do código em python, foi demonstrada ser essencial para o desenvolvimento da pesquisa, pois permitiu a realização da triagem dos medicamentos bem como otimizou o tempo de processamento dos dados. Por outro lado, este código pode ser aprimorado futuramente.

Com a utilização da biblioteca PyMuPDF, é possível extrair o contexto no qual a palavra chave foi encontrada, além de integrar os resultados com a plataforma Power BI, capaz de gerar gráficos em tempo real, a partir de dados coletados com uma ferramenta de *web scraping*, que fará o *download* dos arquivos presentes no site do FDA.

Vale também ressaltar a qualidade dos arquivos disponíveis para a realização da pesquisa, observado a necessidade de melhorara das informações nos bancos de dados tanto dos desenvolvedores como das agências reguladoras, em especial quando se trata da ANVISA.

Por fim, um maior entendimento quanto às técnicas de IA que já foram utilizadas com sucesso, em casos reais de medicamentos aprovados, é a maior contribuição desse trabalho para a comunidade acadêmica em geral, além de expor que a inteligência artificial no processo de desenvolvimento de fármacos não é mais uma promessa, e sim uma realidade.

## REFERÊNCIAS

ANDRADE, C. H. A Inteligência Artificial está mudando a Intuição dos Químicos Medicinais? **XIV Encontro Nacional de Professores de Química Farmacêutica**, Palestra, Universidade Federal de Minas Gerais, agosto de 2023.

BURKI, T. A new paradigm for drug development. **The Lancet Digital Health**, v. 2, n. 5, p. 226–227, Maio 2020.

BARREIRO, E.; ALBERTO, C. **Química Medicinal** - 3.Ed. Artmed Editora, 2014.

CHUN, M. **How Artificial Intelligence is Revolutionizing Drug Discovery**. Disponível em: <<https://blog.petrieflom.law.harvard.edu/2023/03/20/how-artificial-intelligence-is-revolutionizing-drug-discovery/>>. Acesso em: fev. 2023.

CENTER FOR DRUG EVALUATION AND RESEARCH APPLICATION NUMBER: 214787Orig1s000 CLINICAL PHARMACOLOGY REVIEW(S). **FOOD AND DRUG ADMINISTRATION**, 16 jul. 2020. Disponível em: [https://www.accessdata.fda.gov/drugsatfda\\_docs/nda/2020/214787Orig1s000ClinpharmR.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/nda/2020/214787Orig1s000ClinpharmR.pdf). Acesso em: 01 mai 2023

DALAL, D. et al. A systematic overview of bexagliflozin: A type 2 diabetic drug. **Asian Journal of Pharmaceutical Research and Health Care**, v. 15, n. 2, p. 109, 2023.

DARA, S. et al. Machine Learning in Drug Discovery: A Review. **Artificial Intelligence Review**, v. 55, n. 3, 11 Ago 2021.

DANDE, A. A.; PUND, M. A. A Review Study on Applications of Natural Language Processing. **International Journal of Scientific Research in Science, Engineering and Technology**, v. 10, n. 2, p. 122–126, 25 mar. 2023.

DOWDEN, H.; MUNRO, J. Trends in clinical success rates and therapeutic focus. **Nature Reviews Drug Discovery**, v. 18, n. 7, p. 495–496, 8 maio 2019.



FDA (U.S FOOD AND DRUG ADMINISTRATION) **Artificial intelligence in manufacturing**. U.S FOOD AND DRUG ADMINISTRATION, Maryland: FDA, 2023.

FRIEDMAN, L. M.; FURBERG, C.; DEMETS, D. L. **Fundamentals of clinical trials**. New York, Ny Springer New York, 2015.

HALVORSEN, Y. et al. Bexagliflozin as an adjunct to metformin for the treatment of type 2 diabetes in adults: A 24-week, randomized, double-blind, placebo-controlled trial. **Diabetes, Obesity & Metabolism**, v. 25, n. 10, p. 2954–2962, 1 out. 2023.

HOY, S. M. Bexagliflozin: First Approval. **Drugs**, v. 83, n. 5, p. 447–453, 2 mar. 2023

LAMBERTI, M. et al. A Study on the Application and Use of Artificial Intelligence to Support Drug Development. **Clinical Therapeutics**, v. 41, n. 8, p. 1414–1426, Ago 2019.

LAMB, Y. N. Remdesivir: First Approval. **Drugs**, v. 80, p. 1355–1363, 1 set. 2020.

LEVIEN, T. L.; BAKER, D. E. Remdesivir. **Hospital Pharmacy**, v. 58, n. 5, p. 420–430, 6 mar. 2021.

LIANG, G. et al. The emerging roles of artificial intelligence in cancer drug development and precision therapy. **Biomedicine & Pharmacotherapy**, v. 128, n. 110255, p. 110255, 1 ago. 2020.

MUTHUKRISHNAN, N. et al. Brief History of Artificial Intelligence. **Neuroimaging Clinics of North America**, v. 30, n. 4, p. 393–399, 1 nov. 2020.

NUSSINOV, R. et al. AlphaFold, Artificial Intelligence (AI), and Allostery. **The Journal of Physical Chemistry**, v. 126, n. 34, p. 6372–6383, 17 ago. 2022.

NADKARNI, P. M.; OHNO-MACHADO, L.; CHAPMAN, W. W. Natural language processing: an introduction. **Journal of the American Medical Informatics Association**, v. 18, n. 5, p. 544–551, set. 2011.

NÉMETH, R.; KOLTAI, J. Natural language processing: The integration of a new methodological paradigm into sociology. **Intersections**, v. 9, n. 1, p. 5–22, 26 abr. 2023.

PyMuPDF SOFTWARE. Documentation, 2023 – Disponível em: <https://pymupdf.readthedocs.io/en/latest/module.html>. Acesso em outubro de 2023.

PYTHON SOFTWARE FOUNDATION. Python Language Site: Documentation, 2023. Página de documentação. Disponível em <https://docs.python.org/pt-br/3/library/os.html>. Acesso em outubro de 2023.

RASHID, A. Artificial Intelligence Effecting a Paradigm Shift in Drug Development. **SLAS TECHNOLOGY**, v. 26, n. 1, p. 3–15, 17 Set 2020.

SCHLANDER, M. et al. How Much Does It Cost to Research and Develop a New Drug? A Systematic Review and Assessment. **Pharmaco Economics**, v. 39, n. 11, 9 Ago 2021.

SERTKAYA, A. et al. Key cost drivers of pharmaceutical clinical trials in the United States. **Clinical Trials**, v. 13, n. 2, p. 117–126, 8 Fev 2016.

SLIWOSKI, G. et al. Computational Methods in Drug Discovery. **Pharmacological Reviews**, v. 66, n. 1, p. 334–395, 31 dez. 2013.

MINISTÉRIO DA SAÚDE - Secretaria de ciência, tecnologia, inovação e insumos estratégicos. **PROTOCOLO CLÍNICO E DIRETRIZES TERAPÊUTICAS DIABETE MELITO TIPO 2**. Ministério da Saúde, 29 out. 2020. Disponível em: [https://www.gov.br/saude/pt-br/assuntos/protocolos-clinicos-e-diretrizes-terapeuticas-pcdt/arquivos/2020/20201113\\_pcdt\\_diabete\\_melito\\_tipo\\_2\\_29\\_10\\_2020\\_final.pdf](https://www.gov.br/saude/pt-br/assuntos/protocolos-clinicos-e-diretrizes-terapeuticas-pcdt/arquivos/2020/20201113_pcdt_diabete_melito_tipo_2_29_10_2020_final.pdf). Acesso em: 29 jul 2023

SERAFIM, M. et al. Machine learning techniques applied to the drug design and discovery of new antivirals: a brief look over the past decade. **Expert Opinion on Drug Discovery**, v. 16, n. 9, p. 961–975, 7 maio 2021.

TROPSHA, A. Best Practices for QSAR Model Development, Validation, and Exploitation. **Molecular Informatics**, v. 29, n. 6-7, p. 476–488, 6 jul 2010.

U.S. Food And Drugs Administration FDA. Center For Drug Evaluation And Research (org.). **Clinical Pharmacology Review**: Application number: 214787orig1s000. 2020. Disponível em: [https://www.accessdata.fda.gov/drugsatfda\\_docs/nda/2020/214787Orig1s000ClinpharmR.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/nda/2020/214787Orig1s000ClinpharmR.pdf). Acesso em: 1 nov. 2023.

U.S. Food And Drugs Administration FDA. Center For Drug Evaluation And Research (org.). **Integrated Review**: Application number: 214373Orig1s000. 2023. Disponível em: [https://www.accessdata.fda.gov/drugsatfda\\_docs/nda/2023/214373Orig1s000IntegratedR.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/nda/2023/214373Orig1s000IntegratedR.pdf). Acesso em: 1 nov. 2023.

VEKLURY® RENDESIVIR: Pó liofilizado 100mg. Responsável técnico Denise Sunagawa. São Paulo : Gilead Sciences Farmacêutica do Brasil Ltda, 2022. **Bula do profissional**. Disponível em: <https://consultas.anvisa.gov.br/#/bulario/detalhe/1457763?nomeProduto=VEKLURY%C2%AE>. Acesso em: ago. 2023.

WANG, L. et al. Advances of Artificial Intelligence in Anti-Cancer Drug Design: A Review of the Past Decade. **Pharmaceuticals**, v. 16, n. 2, p. 253–253, 7 fev. 2023.

ZHOU, Z. et al. Optimization of Molecules via Deep Reinforcement Learning. **Scientific Reports**, v. 9, n. 1, 24 Jul 2019.