



UNIVERSIDADE D  
COIMBRA

Diogo André Catarino

Relatórios de Estágio e Monografia intitulada “Inteligência Artificial no Desenvolvimento de Novos Fármacos” referentes à Unidade Curricular “Estágio” sob a orientação da Dra. Ana Tomé, da Dra. Patrícia Carrilho e do Professor Doutor Diogo da Fonseca apresentados à Faculdade de Farmácia da Universidade de Coimbra, para apreciação na prestação de provas públicas de Mestrado Integrado em Ciências Farmacêuticas.

Julho de 2020



UNIVERSIDADE D  
COIMBRA

Diogo André Catarino

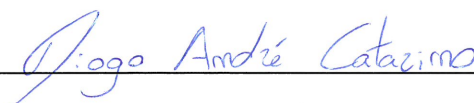
Relatórios de Estágio e Monografia intitulada “Inteligência Artificial no Desenvolvimento de Novos Fármacos” referentes à Unidade Curricular “Estágio” sob a orientação da Dra. Ana Tomé, da Dra. Patrícia Carrilho e do Professor Doutor Diogo da Fonseca, apresentados à Faculdade de Farmácia da Universidade de Coimbra, para apreciação na prestação de provas públicas de Mestrado Integrado em Ciências Farmacêuticas.

Julho de 2020

Eu, Diogo André Catarino, estudante do Mestrado Integrado em Ciências Farmacêuticas, com o nº 2013135051, declaro assumir toda a responsabilidade pelo conteúdo do Documento Relatórios de Estágio e Monografia intitulada “Inteligência Artificial no Desenvolvimento de Novos Fármacos” apresentados à Faculdade de Farmácia da Universidade de Coimbra, no âmbito da unidade de Estágio Curricular.

Mais declaro que este Documento é um trabalho original e que toda e qualquer afirmação ou expressão, por mim utilizada, está referenciada na Bibliografia, segundo os critérios bibliográficos legalmente estabelecidos, salvaguardando sempre os Direitos de Autor, à exceção das minhas opiniões pessoais.

Coimbra, 03 de julho de 2020.



---

(Diogo André Catarino)

## Agradecimentos

---

Aos meus pais, por serem os pilares da minha vida e fazerem de mim a pessoa que hoje sou. É um orgulho para mim ser vosso filho!

Às minhas irmãs, por serem um exemplo de superação, força e trabalho árduo. Obrigado por estarem sempre presentes!

À minha família, pelo incentivo que sempre me deram e continuam a dar.

À minha avó Josefina, onde quer que estejas espero que te sintas orgulhosa de mim.

À Carolina, a quem palavras não chegam para agradecer todo o apoio, incentivo, momentos especiais e amizade! Obrigado por tudo, tornaste esta jornada especial!

À Penas, à Campelo e à Ana Cláudia, a família que Coimbra me deu, obrigado pela amizade e bons momentos que partilhamos juntos. Que este laço que nos une nunca se quebre.

Aos meus amigos, que me acompanharam ao longo desta aventura.

Ao Professor Doutor Diogo da Fonseca, por toda a disponibilidade, preocupação e orientação.

A toda a Equipa Técnica da Farmácia Gonçalves, em particular à Dra. Patrícia Carrilho, pela amizade e ensinamentos que me transmitiram. Foi um privilégio aprender e privar convosco.

À Unidade de Farmacovigilância da Medinfar, em particular a Dra. Ana Tomé, pela forma como me acolheu, pela disponibilidade e pelos conhecimentos que me transmitiu.

A Coimbra, por tantas amizades e boas memórias me ter dado!

### **Parte 1: Relatório de Estágio em Farmacovigilância**

Lista de Abreviaturas.....	8
1. Introdução.....	9
2. Grupo Medinfar .....	10
3. Análise SWOT .....	11
A. Pontos Fortes.....	11
1. Acolhimento e Integração do Estagiário .....	11
2. Plano de Estágio e Acompanhamento do Estagiário .....	11
3. Conhecimentos Adquiridos e Diversidade de Tarefas .....	12
B. Pontos Fracos.....	13
1. Duração do Estágio.....	13
2. Ausência de Contacto com Determinadas Atividades .....	13
C. Oportunidades.....	14
1. Possibilidade de Aprendizagem com Especialistas .....	14
2. Quantidade de Estagiários na UFV .....	14
D. Ameaças.....	14
1. Elevado Número de Duplicados na <i>European Medicines Agency (EMA)</i> .....	14
4. Considerações Finais .....	15
5. Referências Bibliográficas.....	16

### **Parte 2: Relatório de Estágio em Farmácia Comunitária**

Lista de Abreviaturas.....	18
1. Introdução.....	19
2. Análise SWOT .....	20
A. Pontos Fortes.....	20
1. Equipa Técnica .....	20
2. Heterogeneidade de utentes .....	20
3. Determinação dos Parâmetros Bioquímicos.....	20
4. Plano de estudos do MICF .....	21
B. Pontos Fracos.....	21
1. Dermocosmética e Puericultura .....	21
2. Não Preparação de Medicamentos Manipulados .....	22
3. Adaptação à Crise Pandémica .....	22

C. Oportunidades.....	23
1. Formações.....	23
2. Pandemia da COVID-19 .....	23
3. Cartão Municipal Sénior .....	23
D. Ameaças.....	24
1. Inflação de preços durante a pandemia.....	24
2. Concorrência exercida por locais de venda de MNSRM.....	24
3. Caso Prático .....	26
4. Considerações Finais .....	27
5. Referências Bibliográficas.....	28
 <b>Parte 3: Monografia: "Inteligência Artificial no Desenvolvimento de Novos Fármacos"</b>	
Lista de Abreviaturas.....	30
Abstract.....	31
Resumo .....	32
1. Introdução.....	33
2. Descoberta de Novos Fármacos: Paradigma Atual .....	34
3. Inteligência Artificial.....	35
3.1 <i>Machine Learning</i> e <i>Deep Learning</i> .....	36
3.2 Variedade de Algoritmos .....	38
4. Inteligência Artificial na Descoberta de Novos Fármacos .....	40
4.1 Previsão das Propriedades Farmacocinéticas e da Toxicidade de Compostos ..	41
4.2 Utilização em <i>Design De Novo</i> .....	44
4.3 Planeamento de Síntese.....	47
5. Vantagens do Uso de Inteligência Artificial .....	50
6. Desafios na Aplicação da Inteligência Artificial .....	52
7. Conclusão .....	54
8. Referências Bibliográficas.....	55

# Parte I

---

## Relatório de Estágio em Farmacovigilância

Laboratório Medinfar – Produtos Farmacêuticos, SA

Venda Nova



Orientado pela Dra. Ana Tomé

## Lista de Abreviaturas

---

AIM	Autorização de Introdução no Mercado
CEI	Comunidade dos Estados Independentes
EMA	<i>European Medicines Agency</i>
FFUC	Faculdade de Farmácia da Universidade de Coimbra
ICSR	<i>Individual Case Safety Report</i>
MICF	Mestrado Integrado em Ciências Farmacêuticas
MLM	<i>Medical Literature Monitoring</i>
OMS	Organização Mundial de Saúde
PGR	Plano de Gestão de Risco
PSUR	<i>Periodic Safety Update Report</i>
QPPV	<i>Qualified Person Responsible for Pharmacovigilance</i>
RPS	Relatório Periódico de Segurança
SWOT	<i>Strengths Weaknesses Opportunities Threats</i>
UFV	Unidade de Farmacovigilância



## I. Introdução

---

O estágio curricular é uma fase de elevada importância para todos os finalistas do Mestrado Integrado em Ciências Farmacêuticas (MICF), não só porque permite a aplicação de conhecimentos já adquiridos, como também constitui um contacto inicial com a realidade profissional, promovendo uma intensiva aprendizagem quer a nível de novos conhecimentos como de responsabilidade e autonomia. Desta forma, além do estágio obrigatório em farmácia comunitária, a Faculdade de Farmácia da Universidade de Coimbra (FFUC) destaca-se por oferecer aos seus discentes a oportunidade de realizar um segundo estágio curricular em outras áreas de intervenção do farmacêutico, como é o caso da indústria farmacêutica.

O presente relatório diz respeito ao estágio realizado na Unidade de Farmacovigilância (UFV) do Grupo Medinfar, na Venda Nova, sob a orientação da Dra. Ana Tomé, que assume o cargo de Responsável pela Farmacovigilância, também designado *Qualified Person Responsible for Pharmacovigilance* (QPPV). O estágio iniciou a 9 de setembro de 2019 e findou a 29 de novembro de 2019.

A Farmacovigilância é, segundo a Organização Mundial de Saúde (OMS) a ciência e o conjunto de atividades relacionadas com a deteção, avaliação, compreensão e prevenção de reações adversas ou outros problemas relacionados com os medicamentos tendo o objetivo de avaliar a causalidade e gravidade entre a reação adversa e o medicamento, permitindo um aumento da segurança do doente em relação ao uso de medicamentos e a análise do risco-benefício do mesmo (WHO, 2019).

Esta é, portanto, uma ferramenta crucial na indústria farmacêutica para a avaliação do perfil de segurança de um fármaco, uma vez que permite a identificação de novas reações adversas, algumas das quais raras e que dificilmente seriam detetadas nos ensaios clínicos, a observação de um possível aumento da frequência de reações adversas já conhecidas e a descoberta de novas interações medicamentosas.

Este relatório constitui então uma avaliação retrospectiva do estágio recorrendo a uma análise SWOT (*Strengths, Weaknesses, Opportunities, Threats*) onde se abordam os pontos fortes, pontos fracos, oportunidades e ameaças que foram sendo identificados no decorrer do estágio.

## 2. Grupo Medinfar

---

O Grupo Medinfar, fundado em 1970, é um grupo sediado em Lisboa, de capital exclusivamente português e que atua em diversas áreas, sendo especializado em investigação e desenvolvimento, fabrico de produtos farmacêuticos, de dermocosméticos e de suplementos, distribuição e comercialização (GRUPO MEDINFAR, 2019).

Não se cingindo apenas às suas próprias marcas, a Medinfar também intervém na produção e comercialização de produtos licenciados em parceria com algumas das maiores empresas farmacêuticas do panorama mundial (GRUPO MEDINFAR, 2019).

O grupo é constituído por diversos sectores, nomeadamente pela Farma (medicamentos de prescrição relativos às áreas respiratórias, dermatologia, diabetes e obesidade), a GP Genéricos Portugueses ( com foco em medicamentos genéricos), *Consumer Health* (medicamentos não sujeitos a receita médica e outros produtos de saúde), Farmalabor (unidade industrial e de produção), Medinfar Sorológico (fármacos veterinários, equipamentos e vacinas de rebanho) e DVINE (marca de cosmética portuguesa com foco no anti envelhecimento).

Sendo detentor de um vasto portefólio de produtos, o grupo atua em diversas áreas da indústria farmacêutica, tais como: Dermatologia, Antibióticos, Antifúngicos, Urologia, Otorrinolaringologia, Osteoporose, Cardiologia, Imunologia, entre outras.

Graças a uma crescente internacionalização com vista a promover o crescimento do grupo, a Medinfar encontra-se atualmente presente em mais de 50 países pertencentes a diferentes regiões do globo, incluindo a Europa, a África Francófona, o Médio Oriente, a Comunidade dos Estados Independentes (CEI) e a Ásia, destacando também a presença de uma filial em Marrocos (GRUPO MEDINFAR, 2019).

### 3. Análise SWOT

---

#### A. Pontos Fortes

##### 1. Acolhimento e Integração do Estagiário

O primeiro dia de estágio foi reservado não só à integração dos novos estagiários de diferentes áreas bem como ao acolhimento de novos colaboradores do grupo. Após uma visita guiada às instalações da Medinfar, o restante “*Welcome Day*” foi reservado a diversas apresentações, realizadas por diferentes colaboradores, que abordavam temas que iam desde a apresentação do Grupo Medinfar, processos de segurança da empresa, código deontológico da Medinfar e passando também por uma breve introdução à Farmacovigilância.

Este dia de acolhimento foi, para mim, muito importante porque não só me forneceu novos conhecimentos que vieram a ser necessários ao longo do meu estágio, como permitiu também um contacto com diferentes colaboradores que sempre se mostraram simpáticos e disponíveis para ajudar, o que me deixou mais tranquilo e bem integrado num espaço que me era totalmente desconhecido até à data e demonstrou o bom ambiente existente no Grupo.

##### 2. Plano de Estágio e Acompanhamento do Estagiário

Foi no segundo dia que conheci o meu local de trabalho e tive um maior contacto com a minha orientadora, a Dra. Ana Tomé, que me apresentou aquele que seria o meu plano de estágio. Inicialmente iria ter uma componente mais teórica onde me iria focar em ler as Boas Práticas de Farmacovigilância e os Procedimentos Operativos Normalizados relativos à UFV com o intuito de me contextualizar com a forma como eram realizadas as variadas tarefas da Unidade. Só depois entraria numa componente mais prática do estágio onde, de forma gradual a nível de exigência e sempre com a supervisão adequada, fui desempenhando diversas funções das quais saliento: a elaboração de Planos de Gestão de Risco (PGRs), que permitem fazer uma análise do perfil de segurança de um medicamento e indicam medidas de minimização do risco; a atuação na base de dados *PcVmanager*, onde ao longo do estágio fui registando os casos de reações adversas, também designados *Individual Case Safety Report (ICSR)*, oriundos de diferentes fontes (espontâneas, literatura ou *EudraVigilance*); e a deteção de ICSR duplicados no sistema.

É importante referir que, ao longo do estágio, foram realizadas reuniões de acompanhamento com o objetivo não só de saberem a minha opinião sobre as atividades que tinha vindo a desempenhar, mas também para avaliarem a minha prestação, propondo-me sempre novos desafios a cada reunião.

A presença de um plano de estágio estruturado, o facto de ter um espaço de trabalho exclusivamente meu e a realização de reuniões de acompanhamento demonstraram de forma evidente a atenção e preocupação que o Grupo Medinfar, nomeadamente a Dra. Ana Tomé, teve relativamente à minha aprendizagem, algo que me motivou imenso na realização das minhas tarefas e a contribuir da melhor maneira para o desempenho da empresa.

### 3. Conhecimentos Adquiridos e Diversidade de Tarefas

Durante o estágio, sempre com a supervisão e ajuda adequada da Dra. Ana Tomé, foram-me propostas diversas tarefas e desafios que muito contribuíram para uma maior aprendizagem e produtividade da minha parte e cultivaram em mim um sentido de responsabilidade que sem dúvida fará de mim um melhor profissional de saúde no futuro.

Inicialmente, comecei por me familiarizar com a base de dados da empresa onde os casos de reações adversas eram processados e arquivados, o *PcVmanager*. Pude inclusive assistir a uma formação sobre a utilização desta plataforma que foi muito útil no decorrer do meu percurso na Medinfar. As minhas tarefas no *PcVmanager* englobavam a deteção e correção de casos duplicados, a triagem e introdução de casos oriundos do *EudraVigilance* e a introdução de casos espontâneos ou de literatura. Realizei também o preenchimento de folhas de rosto onde se documentava a avaliação e triagem de casos relativos a medicamentos de uso humano e cosméticos.

Tive também contacto com a pesquisa no *medical literature monitoring (MLM) service*, através do *EudraVigilance*, sendo esta realizada por substância ativa e em países onde uma das empresas constituintes do Grupo Medinfar possua a titularidade de Autorização de Introdução no Mercado (AIM).

Já numa fase mais terminal do estágio, fui incluído na elaboração de um Plano de Gestão de Risco. Esta foi sem dúvida a tarefa mais exigente para mim, embora ao mesmo tempo tenha sido a que me deu maior motivação e satisfação em realizar. A elaboração de um PGR envolve, em primeiro lugar, a consulta das Boas Práticas de Farmacovigilância de forma a identificar os requisitos aplicáveis na elaboração do mesmo, requisitos esses que variam mediante a classificação do medicamento (genérico, uso bem estabelecido, híbrido, etc.). Após esta análise, é importante fazer uma recolha de dados que envolve a pesquisa em bases de dados, na literatura e nos documentos de referência do medicamento – Resumo das Características do Medicamento e Folheto Informativo. Desta forma, é possível estabelecer o perfil de segurança de um determinado medicamento, expondo os riscos identificados e potenciais relativos ao seu uso e identificando as populações onde existe falta de informação. Por fim, o PGR apresenta uma série de medidas de minimização de risco (incluindo uma avaliação da

efetividade das mesmas) de forma a melhorar a utilização do medicamento e a promover a segurança do doente.

Em suma, embora não tenha abordado todas as tarefas possíveis de realizar relativamente à Farmacovigilância, todas as atividades acima referidas permitiram que o meu estágio fosse dinâmico e desafiante, promovendo uma aprendizagem enriquecedora e aperfeiçoando a minha capacidade de trabalho.

## **B. Pontos Fracos**

### I. Duração do Estágio

A duração constitui para mim um aspeto limitante num estágio em indústria farmacêutica. O enorme volume de nova informação adquirida faz com que o tempo necessário até se atingir um grau de autonomia aceitável seja elevado, sendo que quando esta autonomia é atingida, o estágio já se encontra muitas vezes numa fase terminal, impossibilitando uma aplicação plena dos nossos conhecimentos.

Para além desta limitação, a curta duração de estágio impossibilita muitas vezes a realização de algumas tarefas que não são tão regularmente executadas ou que demoram mais tempo a serem realizadas. No meu caso, o término do estágio não me permitiu completar da forma pretendida um processo de gestão de sinal para um determinado medicamento, algo que poderia ter sido relevante para a minha aprendizagem.

### 2. Ausência de Contacto com Determinadas Atividades

Tal como referido no ponto anterior, a duração do estágio muitas vezes impossibilita o contacto com determinadas áreas e/ou atividades. No meu caso em particular, considero um ponto fraco a ausência de contacto com toda a área dos medicamentos veterinários, uma área deveras interessante e desconhecida para mim.

Também não pude participar na elaboração de um Relatório Periódico de Segurança (RPS), também designado por *Periodic Safety Update Report* (PSUR). No entanto, isto deveu-se ao facto de os medicamentos que constituem o portefólio da Medinfar terem um período relativamente largo entre cada submissão de RPS e de o meu estágio não ter abrangido nenhum desses períodos.

Concluindo, e realçando os muitos conhecimentos adquiridos noutras atividades de Farmacovigilância, considero que um contacto com a área veterinária e a elaboração de um RPS teriam completado substancialmente a minha formação na área.

## **C. Oportunidades**

### I. Possibilidade de Aprendizagem com Especialistas

Os estágios curriculares permitem aos estagiários o contacto com especialistas nas mais diversas áreas farmacêuticas. Tal como já referi anteriormente, a Dra. Ana Tomé, uma profissional com vasto conhecimento e experiência em Farmacovigilância, ficou a cargo da minha orientação. Ter a oportunidade de aprender com ela foi sem dúvida algo que promoveu a minha aprendizagem e capacidade de trabalho, uma vez que a sua ajuda permitiu uma agilização de algumas das tarefas que executei e que, sem essa ajuda, teriam sido muito mais complexas.

### 2. Quantidade de Estagiários na UFV

O facto de ter sido o único estagiário na UFV durante o período de setembro a novembro foi uma ótima oportunidade de aprendizagem porque permitiu que mais tarefas fossem destacadas para mim e que sempre que eu precisasse de qualquer tipo de ajuda ou orientação, esta fosse fornecida de forma mais rápida. Desta forma, senti sempre ao longo do meu percurso que recebi todo o acompanhamento necessário para a minha progressão.

## **D. Ameaças**

### I. Elevado Número de Duplicados na *European Medicines Agency (EMA)*

Os duplicados constituem casos que se referem à mesma reação adversa, ocorrida no mesmo doente e referente ao mesmo medicamento. Como todas as UFV dos diferentes titulares de AIM realizam o mesmo tipo de monitorização de literatura e substâncias ativas, é comum que muitas vezes os mesmos casos sejam submetidos por diferentes fontes ao *Eudravigilance*, o que promove, a meu ver, um aumento de entropia no sistema e origina um uso desnecessário de recursos humanos e materiais uma vez que é necessário fazer deteção destes mesmos duplicados.

Atualmente, a EMA reúne todos estes duplicados e funde-os todos num caso único, designando o por *Master Case*. No entanto, na minha opinião, outras medidas deviam ser criadas para diminuir a submissão excessiva dos mesmos casos.

#### 4. Considerações Finais

---

Este estágio curricular foi, para mim, uma oportunidade única de adquirir conhecimentos e de conhecer um pouco o mundo que é a Farmacovigilância. Após uma análise retrospectiva, facilmente percebo que houve um aumento substancial da minha responsabilidade, autonomia, exigência, atenção aos detalhes e capacidade de trabalho, qualidades essas fundamentais para o meu futuro como profissional de saúde.

Apreendi que a Farmacovigilância é muito mais do que a simples recolha de reações adversas. É um pilar que promove um uso mais seguro do medicamento, avalia o risco-benefício do mesmo, protege a segurança do doente e, muito importante, permite uma ação proativa dos titulares de AIM através de medidas de minimização de risco. Sendo assim, na minha opinião, esta será no futuro uma área de grande relevância e com um maior peso na indústria farmacêutica.

Por fim, resta-me agradecer ao Grupo Medinfar, em particular à Dra. Ana Tomé, por todos dos desafios, ensinamentos e disponibilidade e acima de tudo por terem promovido em mim o interesse por esta área tão importante que é a Farmacovigilância.

## 5. Referências Bibliográficas

---

GRUPO MEDINFAR – Quem Somos. [Consultado a 14 de dezembro de 2019]. Disponível na Internet: <https://www.medinfar.pt/pt/sobre-nos/quem-somos>

WORLD HEALTH ORGANIZATION (WHO) – Pharmacovigilance. [Consultado a 14 de dezembro de 2019]. Disponível na Internet: [https://www.who.int/medicines/areas/quality\\_safety/safety\\_efficacy/pharmvigi/en/](https://www.who.int/medicines/areas/quality_safety/safety_efficacy/pharmvigi/en/)



# Parte 2

---

## Relatório de Estágio em Farmácia Comunitária

Farmácia Gonçalves

Odivelas



farmácia  
gonçalves

Orientado pela Dra. Patrícia Carrilho

## Lista de Abreviaturas

---

EPI	Equipamento de Proteção Individual
FFUC	Faculdade de Farmácia da Universidade de Coimbra
INR	Razão Normalizada Internacional
MICF	Mestrado Integrado em Ciências Farmacêuticas
MNSRM	Medicamento Não Sujeito a Receita Médica
MSRM	Medicamento Sujeito a Receita Médica
SNS	Serviço Nacional de Saúde
SWOT	<i>Strengths Weaknesses Opportunities Threats</i>

## I. Introdução

---

No âmbito da unidade curricular “Estágio Curricular”, o seguinte relatório foi elaborado após o meu estágio na Farmácia Gonçalves, em Odivelas, sob a orientação da Dr.<sup>a</sup> Patrícia Carrilho.

As farmácias comunitárias são fulcrais para garantir os pilares preconizados no Serviço Nacional de Saúde (SNS): acessibilidade ao medicamento e equidade na prestação de cuidados de saúde de qualidade a todos os cidadãos independentemente da sua localização geográfica (ORDEM DOS FARMACÊUTICOS, 2020a). Isto é possível, não só devido à vasta distribuição geográfica das farmácias e fácil acessibilidade por parte das comunidades, mas também pelos seus profissionais, que possuem elevadas competências técnico-científicas.

Desta forma, este é o espaço mais adequado para que o farmacêutico execute, de acordo com o que é evidenciado no artigo 9º do Código Deontológico da Ordem dos Farmacêuticos, “todas as tarefas que ao medicamento concernem, todas as que respeitam às análises clínicas ou análises de outra natureza (...) e todas as ações de educação dirigidas à comunidade no âmbito da promoção da saúde” (ORDEM DOS FARMACÊUTICOS, 2020b).

Sendo assim, após a formação teórica e multidisciplinar que a Faculdade de Farmácia da Universidade de Coimbra (FFUC) me forneceu, este estágio constituiu um desafio prático que me permitiu não só aplicar e aprimorar os conhecimentos já adquiridos, mas também perceber a importância que o farmacêutico tem para a sociedade, papel esse tantas vezes desvalorizado mas que é essencial para a promoção do uso racional do medicamento, incentivo de estilos de vida mais saudáveis, aumento da literacia em saúde e uma melhor adesão à terapêutica.

Serve este relatório como uma análise retrospectiva do estágio realizado na Farmácia Gonçalves, aplicando um modelo de avaliação SWOT (Strengths, Weaknesses, Opportunities, Threats) que permite avaliar os pontos fortes e fracos, as oportunidades e as ameaças identificadas ao longo deste desafio.

## 2. Análise SWOT

---

### A. Pontos Fortes

#### 1. Equipa Técnica

A equipa técnica da Farmácia Gonçalves foi, sem dúvida alguma, crucial para a minha aprendizagem e para o sucesso do estágio. A presença de profissionais de excelência, quer a nível profissional quer a nível pessoal, permitiu que a minha integração na equipa fosse mais fácil e que desde cedo me sentisse à vontade num ambiente que era totalmente novo para mim.

Rapidamente percebi que o principal foco da farmácia era o utente e o seu bem-estar, e a observação das práticas e aconselhamentos realizados por esta equipa permitiu o meu crescimento a nível profissional, aumentando os meus conhecimentos sobre questões e situações mais comuns no quotidiano da farmácia e melhorando a minha capacidade de comunicação.

Desta forma, o contacto com os elementos desta equipa, que sempre se mostraram disponíveis para me ajudar e interessados em elucidar-me sobre as mais variadas questões, permitiu-me ganhar confiança, tranquilidade e segurança em todas as tarefas que desempenhei, especialmente no atendimento ao público e no aconselhamento do utente.

#### 2. Heterogeneidade de utentes

A maioria dos utentes da farmácia é constituída por clientes fidelizados locais, havendo a tendência a que pertençam a uma população sénior. No entanto, a proximidade para com o metropolitano e outros meios de transporte urbanos, permite uma elevada afluência de utentes de ocasião das mais variadas faixas etárias.

Sendo assim, esta heterogeneidade de utentes possibilitou-me o contacto com uma grande variedade de atendimentos e uma melhor aplicação dos meus conhecimentos, oferecendo a oportunidade de contactar com vasta parte dos serviços e produtos existentes numa farmácia comunitária, tendo-me sido possível efetuara dispensa e aconselhamento de medicamentos sujeitos a receita médica (MSRM), de medicamentos não sujeitos a receita médica (MNSRM), de produtos cosméticos, suplementos alimentares, produtos homeopáticos e produtos ortopédicos, entre outros.

#### 3. Determinação dos Parâmetros Bioquímicos

A possibilidade de poder realizar a determinação de diversos parâmetros, entre os quais a glicémia, os triglicéridos, o colesterol, o INR e a medição da pressão arterial não só

constitui uma atividade gratificante a nível pessoal, como também permite a criação de um elo de maior confiança entre utente e farmacêutico.

O facto de na Farmácia Gonçalves existir um gabinete privado e específico para esta atividade, longe da agitação que se verifica por vezes ao balcão, torna a comunicação com o utente mais eficaz e tranquila, permitindo perceber quais as dúvidas e estilo de vida do mesmo e criando a oportunidade para o aconselhar, facilitando a correção de possíveis situações de risco e promovendo a adesão à terapêutica.

Esta atividade é, portanto, um ponto forte do meu estágio porque me fez perceber que na maioria das vezes é na farmácia que a população em geral procura obter informação e aconselhamento, reforçando assim a importância do farmacêutico como profissional de saúde de primeira linha.

#### 4. Plano de estudos do MICE

O plano de estudos apresentado pelo MICE fornece uma base teórica abrangente e diversificada, como forma de preparar o aluno para qualquer uma das áreas farmacêuticas existentes no mercado de trabalho. Relativamente à área da Farmácia Comunitária, há uma clara preocupação em preparar os estagiários para a realidade do aconselhamento ao utente, algo que é visível nos conhecimentos adquiridos em unidades curriculares como Farmacologia ou Indicação Farmacêutica. Ao longo do estágio, senti efetivamente que essa formação teórica foi uma mais valia para o meu desempenho.

No entanto, o aconselhamento de produtos de outras áreas, das quais realço a dermocosmética, a puericultura e os medicamentos de uso veterinário, constituiu um desafio para mim, de tal forma que necessitei de recorrer à ajuda da equipa da Farmácia Gonçalves em quase todas as situações. Sendo assim, seria ideal que as unidades curriculares relativas a estas áreas não se focassem apenas em fornecer conhecimentos teóricos sobre os produtos, mas também tentarem promover atividades alinhadas com a realidade prática de farmácia comunitária.

### **B. Pontos Fracos**

#### I. Dermocosmética e Puericultura

Apesar da Farmácia Gonçalves dispor de um variado conjunto de produtos de dermocosmética e puericultura à disposição do utente, esta foi uma área com a qual não tive muito contacto. Em retrospectiva, considero que o facto de existir um local de atendimento específico para este tipo de situações fez com que o meu contacto com atendimentos deste

âmbito fosse mais limitado, impedindo o desenvolvimento das minhas capacidades nesta vertente.

No entanto, é importante realçar que quer os elementos que trabalhavam exclusivamente na parte da dermocosmética e puericultura, quer os restantes profissionais da equipa procuraram sempre esclarecer as minhas dúvidas sobre este tipo de produtos.

## 2. Não Preparação de Medicamentos Manipulados

Os medicamentos são considerados manipulados quando estes são preparados segundo uma farmacopeia ou formulário, designados por Preparados Oficiais, ou de acordo com uma receita médica que especifique o doente alvo do medicamento, conhecidas por Fórmulas Magistrais (INFARMED, I.P.). Desta forma, é possível preparar medicamentos adaptados para determinadas situações ou patologias quando as posologias ou preparações disponíveis no mercado não são as ideais. Esta é uma situação muito comum no caso do uso pediátrico de medicamentos.

Embora a Farmácia Gonçalves possua instalações e equipamentos para a preparação de medicamentos manipulados, a pouca procura por este tipo de preparação fez com que a minha experiência neste âmbito tenha sido em parte limitada, restringindo-se à simulação do preenchimento de uma ficha de preparação de medicamentos manipulados.

## 3. Adaptação à Crise Pandémica

Durante o decorrer do estágio, o país e o mundo encontraram-se num ambiente de pandemia devido a um novo coronavírus (SARS-CoV-2), também designado por COVID-19. Esta foi uma fase bastante complexa para as farmácias, onde o número de utentes aumentou exponencialmente devido ao medo de uma possível rutura de stocks, enquanto era necessário aplicar diversas medidas preventivas para minimizar a propagação do vírus.

Face a isto, a farmácia adotou inúmeras medidas que promoviam a proteção tanto dos elementos da sua equipa como dos utentes da farmácia, entre as quais a utilização de EPI adequado (bata, máscara, luvas, viseira), higienização do balcão após cada atendimento, disponibilização de soluções antissépticas alcoólicas e limitação do número de pessoas a entrar dentro da farmácia. No entanto, a meu ver, houve outras medidas que não foram consideradas e que considero como ponto fraco a não implementação das mesmas. A primeira, prendeu-se com o facto de a equipa ter continuado a trabalhar toda junta, isto é, sem se criarem turnos específicos de forma a que a possível infeção de um dos elementos não originasse o isolamento de toda a equipa. A outra, é referente ao facto de os seis balcões disponíveis para atendimento

terem continuado todos abertos, o que impossibilitava um distanciamento eficaz entre os profissionais da equipa técnica.

Felizmente, nenhum colaborador da farmácia foi infetado pelo vírus durante o decorrer do estágio, mas considero que a não implementação das medidas acima referidas constituiu um risco desnecessário para todos e que podia originar o fecho temporário da farmácia caso a equipa técnica tivesse de entrar em isolamento.

## **C. Oportunidades**

### I. Formações

Ao longo do estágio, tive oportunidade de contactar com delegados de informação de laboratórios e indústrias farmacêuticas que me forneceram diversas formações sobre os seus produtos. Tendo em conta que o mercado farmacêutico está em constante movimento, com o surgimento de novos produtos para diversas patologias e situações, é muito importante que o farmacêutico procure estar a par de todas estas novidades de forma a zelar pelo bem-estar e saúde do utente.

Por esse motivo, considero que estas formações foram uma mais valia na minha formação porque não só permitiram enriquecer os meus conhecimentos sobre determinados MNSRM que tínhamos à nossa disposição, como também melhorou a minha capacidade de aconselhamento para com o utente.

### 2. Pandemia da COVID-19

Devido à pandemia do vírus SARS-CoV-2 houve, nas primeiras semanas, um enorme fluxo de utentes na farmácia, provocando assim um aumento da carga de trabalho que toda a equipa técnica teve de suportar.

Face a esta situação negativa, tive a necessidade de me adaptar a esta azáfama e, com a ajuda dos restantes profissionais que sempre se mostraram disponíveis sempre que precisei, tornei-me mais pró-ativo e responsável. Em retrospectiva, considero que esta situação promoveu um maior ganho de experiência, conhecimentos, e capacidade de trabalho, fomentando o meu amadurecimento como futuro profissional de saúde.

### 3. Cartão Municipal Sénior

Como já referido anteriormente, a maior parte dos utentes da Farmácia Gonçalves é constituída por uma população sénior residente no município de Odivelas. Como forma de melhorar a qualidade de vida dos seus munícipes, a Câmara Municipal de Odivelas criou um cartão, de carácter pessoal e intransmissível, que de forma gratuita concede benefícios a todos

os interessados com idade igual ou superior a 60 anos (CÂMARA MUNICIPAL DE ODIVELAS).

Na realidade da Farmácia Gonçalves, este cartão beneficia o utente com um desconto de 12% sobre o valor total a pagar quer no receituário participado, quer no receituário não participado ou nas situações que não envolvem receituário de todo (CÂMARA MUNICIPAL DE ODIVELAS).

Considero isto uma oportunidade porque, ao longo do meu estágio, verifiquei que esta estratégia permitiu atrair novos utentes à farmácia que, após o atendimento e aconselhamento recebido por parte da equipa técnica, se tornaram novos clientes fidelizados da farmácia.

## **D. Ameaças**

### 1. Inflação de preços durante a pandemia

Como seria de esperar, a pandemia que se viveu durante o período de estágio levou a um incremento na procura de determinados produtos como máscaras, luvas e soluções antissépticas alcoólicas. No entanto, esta procura excessiva levou, inicialmente, a uma rutura de stocks, após a qual houve um aumento excessivo nos preços dos produtos supracitados.

A falta de controlo nos preços praticados pelos fornecedores, fez com que as farmácias ficassem de mãos atadas e, ao comprarem os produtos por preços muito mais elevados viram-se obrigadas a vender a preços igualmente altos.

Considero que isto foi uma ameaça ao meu estágio e que foi uma situação que se refletiu negativamente na imagem das farmácias, pelo que deveria existir um maior controlo dos preços praticados pelos fornecedores.

### 2. Concorrência exercida por locais de venda de MNSRM

As grandes cadeias de locais de venda de MNSRM, devido ao grande volume de unidades que adquirem, conseguem introduzir no mercado MNSRM a preços com os quais as farmácias comunitárias não conseguem competir, mesmo quando estas pertencem a grupos de farmácias.

Esta situação induz um outro problema, a preferência do utente em adquirir os produtos a preços mais baixos, ignorando ao mesmo tempo algo tão importante como o aconselhamento farmacêutico. Por conseguinte, há também uma promoção da automedicação na população, o que pode constituir um risco para a saúde pública.

Embora no caso da Farmácia Gonçalves, que possui um grande volume de utentes fidelizados, este não seja um problema de grande relevo, considero esta situação uma ameaça



para o panorama geral das farmácias comunitárias, que terão de tirar partido dos conhecimentos e competências das suas equipas para promoverem uma maior fidelização dos utentes e um adequado uso do medicamento.

### 3. Caso Prático

---

Uma utente do sexo feminino, com cerca de 25 anos, dirige-se à farmácia e refere que tem episódios de tosse seca e irritativa e solicita algo que a alivie, referindo que já tentou tomar dois tipos de xarope, um exclusivamente para tosse seca e outro com ação tanto para tosse seca como com expetoração, e que nenhum resolveu a sua situação. Após algumas questões, a utente explica que estes episódios têm tendência a ocorrer à noite e que para além da tosse, possui também irritação ocular.

Com base nestas informações, e em vez de um xarope, sugeri a toma de um anti-histamínico sob suspeita de ser um caso de tosse alérgica. Desta forma, aconselhei a toma de comprimidos de cetirizina, um antagonista seletivo dos recetores H1 periféricos, na posologia recomendada para adultos, um comprimido de 10 mg uma vez por dia, preferencialmente à noite. Adicionalmente, propus a toma de pastilhas para a garganta, mais concretamente as Strepils Tosse<sup>®</sup>, como forma de aliviar a garganta irritada devido à tosse.

#### 4. Considerações Finais

---

Embora tivesse algum receio relativamente a este estágio, devido à necessidade de comunicar com os utentes e à responsabilidade que é necessária para os aconselhar da melhor forma, a presença de uma equipa profissional, competente e preocupada com o meu desenvolvimento, permitiu que perdesse alguma da insegurança que tinha inicialmente e que ganhasse confiança nas atividades que desempenhei ao longo do estágio.

Completo este estágio sem ter qualquer dúvida sobre a importância do farmacêutico como profissional de saúde que, realizando as suas tarefas de forma eticamente correta e responsável, constitui uma peça essencial para a saúde e bem-estar da população, com quem tem uma boa relação de confiança e proximidade.

Esta foi sem dúvida uma experiência gratificante e enriquecedora, que promoveu o meu crescimento tanto a nível pessoal, através do contacto diário com os utentes que nos procuravam com o intuito de esclarecer as suas dúvidas, como a nível profissional, onde pude aplicar todos os conhecimentos teóricos adquiridos no MICF e assim ter uma perceção mais real das atividades desempenhadas pelo farmacêutico no âmbito da farmácia comunitária.

Por fim, resta-me agradecer à Farmácia Gonçalves e a todos os seus profissionais pelos ensinamentos que me transmitiram, por toda a disponibilidade, pela amizade e por terem fomentado em mim os valores de que o utente é sempre a principal preocupação do farmacêutico.

## 5. Referências Bibliográficas

---

CÂMARA MUNICIPAL DE ODIVELAS – Cartão Municipal Sénior. [Consultado a 16 de junho de 2020]. Disponível na Internet: <https://www.cm-odivelas.pt/areas-de-intervencao/intervencao-social/cartao-municipal-senior>

INFARMED, I.P. – Medicamentos Manipulados. [Consultado a 16 de junho de 2020]. Disponível na Internet: <https://www.infarmed.pt/web/infarmed/entidades/medicamentos-uso-humano/inspecao-medicamentos/medicamentos-manipulados>

ORDEM DOS FARMACÊUTICOS (2020a) – A Farmácia Comunitária. [Consultado a 16 de junho de 2020]. Disponível na Internet: <https://www.ordemfarmaceuticos.pt/pt/areas-profissionais/farmacia-comunitaria/a-farmacia-comunitaria/>

ORDEM DOS FARMACÊUTICOS (2020b) – Código Deontológico da Ordem dos Farmacêuticos. [Consultado a 16 de junho de 2020]. Disponível na Internet: [https://www.ordemfarmaceuticos.pt/fotos/documentos/codigo\\_deontologico\\_da\\_of\\_4436676175988472c14020.pdf](https://www.ordemfarmaceuticos.pt/fotos/documentos/codigo_deontologico_da_of_4436676175988472c14020.pdf)

# Parte 3

---

Monografia

## “Inteligência Artificial no Desenvolvimento de Novos Fármacos”



UNIVERSIDADE D  
COIMBRA

Orientada pelo Professor Doutor Diogo André Afonso da Fonseca

## Lista de Abreviaturas

---

ADME	Absorção, Distribuição, Metabolismo e Excreção
AECF	<i>Adaptative Ensemble Classification Framework</i>
ANN	<i>Artificial Neural Network</i>
DL	<i>Deep Learning</i>
DNN	<i>Deep Neural Network</i>
GENTRL	<i>Generative Tensorial Reinforcement Learning</i>
HTS	<i>High-Throughput Screening</i>
IA	Inteligência Artificial
KNN	<i>K-Nearest Neighbor</i>
ML	<i>Machine Learning</i>
PPAR	<i>Peroxisome Proliferator-Activated Receptor</i>
RBES	<i>Rule-Based Expert System</i>
RNN	<i>Recurrent Neural Network</i>
SMILES	<i>Simplified Molecular Input Line Entry System</i>
SVM	<i>Support Vector Machine</i>
VAE	<i>Variational Autoencoders</i>

## Abstract

---

Nowadays, the discovery and development of new drugs proves to be increasingly unable to produce satisfactory results, becoming an increasingly time consuming and expensive process for the pharmaceutical industries.

In this way, it is necessary to find a strategy that allows to stop this negative trend and the renewed interest in artificial intelligence might be a possible solution for the problem. This technology, which focuses on the demonstration of intelligence from computers, has a high capacity to process data and make predictions, being a potential added value for the pharmaceutical area, making processes faster, cheaper and more effective.

Although its application can occur in all phases of the development process, this report will only cover its use in three specific areas: prediction of pharmacokinetic properties and toxicity of compounds, use in de novo design and synthesis planning.

**Keywords:** Artificial intelligence, Machine Learning, Deep Learning, Design de novo, Synthesis planning.

## Resumo

---

Atualmente, o processo de descoberta e desenvolvimento de novos fármacos mostra-se cada vez mais incapaz de produzir resultados satisfatórios, tornando-se cada vez mais demorado e dispendioso para as indústrias farmacêuticas.

Desta forma, é necessário encontrar uma estratégia que permita travar esta tendência negativa sendo o renovado interesse na inteligência artificial uma possível solução para o problema. Esta tecnologia, que se foca na demonstração de inteligência por parte de computadores, possui uma elevada capacidade de processar dados e de fazer previsões sendo uma potencial mais valia para a área farmacêutica, tornando os processos mais rápidos, baratos e eficazes.

Embora a sua aplicação possa ocorrer em todas as fases do processo de desenvolvimento, nesta monografia será abordada apenas a sua utilização em três áreas específicas: previsão de propriedades farmacocinéticas e da toxicidade de compostos, utilização em *design de novo* e planeamento de síntese.

**Palavras-chave:** Inteligência Artificial, *Machine Learning*, *Deep Learning*, *Design de novo*, Planeamento de Síntese.



## I. Introdução

---

A descoberta de novos fármacos é uma área multidisciplinar que reúne o contributo de diversas ciências como a farmacologia, a bioquímica, a química analítica, a microbiologia e inúmeras outras, que contribuíram e continuam a contribuir para o atual processo de desenvolvimento de novas substâncias farmacológicas seguras e eficazes (DREWS, 2000).

Desde a antiguidade que se utilizam substâncias farmacologicamente ativas obtidas a partir de produtos naturais mas foi apenas no início do século passado que se denotou uma maior evolução na procura de novos compostos, facto que se refletiu e, ao mesmo tempo, só se tornou possível, devido aos muitos avanços científicos que marcaram o século XX, nomeadamente o desenvolvimento da química analítica, que possibilitou o isolamento de substâncias químicas, os testes *in vivo* em modelos animais e, mais recentemente, as tecnologias de *screening* como o *High-Throughput Screening* ou HTS (PINA *et al.*, 2010).

Não obstante todos os avanços tecnológicos e inovações concretizadas, trazer novos fármacos para o mercado continua a ser um processo moroso, extremamente caro e com uma taxa de insucesso elevada, especialmente nas fases finais do desenvolvimento, em que já foi investido muito dinheiro, originando elevados prejuízos para as indústrias farmacêuticas (ZURDO, 2013). Esta situação é visível considerando que em 2000 as despesas associadas à investigação e desenvolvimento farmacêutico na Europa rondavam os 18 milhões de euros, valores esses que tiveram um incremento para cerca de 35 milhões em 2017 (EFPIA, 2019). Contrariamente a esta subida dos custos de investigação, o número de novas entidades químicas a serem aprovadas na Europa manteve-se relativamente constante, uma vez que, entre 1999 e 2003, foram aprovadas 62 novas moléculas e, comparativamente, no período entre 2014 e 2018 o valor de aprovações subiu apenas para 67 (EFPIA, 2019), o que demonstra que o aumento da despesa com Investigação e Desenvolvimento não tem sido acompanhado por um aumento dos resultados equivalente.

Nesta monografia, será feita inicialmente uma contextualização do paradigma atual da descoberta de fármacos seguida de uma elucidação sobre alguns conceitos base associados à Inteligência Artificial (IA). Posteriormente, o foco passará a ser a potencial utilização desta tecnologia em algumas das fases do desenvolvimento de fármacos, seguida de uma análise das vantagens dessa utilização e os desafios associados à sua implementação.

## 2. Descoberta de Novos Fármacos: Paradigma Atual

---

Estima-se que o processo de descoberta e desenvolvimento de novos fármacos demore cerca de 14 anos para se concretizar, estando associados investimentos na ordem dos 800 milhões de dólares (BATOOL *et al.*, 2019).

Nos últimos anos, o passo inicial para o começo deste processo envolve a identificação de alvos moleculares que apresentem uma possível relação com uma determinada patologia (BARNASH *et al.*, 2017). Com base nesses alvos, e recorrendo na maioria das vezes a técnicas de HTS, são analisadas bibliotecas de compostos com o intuito de localizar compostos *hit*, moléculas que demonstram capacidade de ligação aos primeiros (BLEICHER *et al.*, 2003). Estes compostos são, posteriormente, refinados com o intuito de os tornar mais potentes, mais seletivos e com propriedades farmacocinéticas adequadas, tornando-se compostos *lead* (HUGHES *et al.*, 2011).

Antes destas moléculas *lead* avançarem para os ensaios clínicos como candidatos a fármaco, é necessária uma nova otimização como forma de não só manter as propriedades já favoráveis, mas também para colmatar possíveis falhas que possam existir na sua estrutura, sendo para isso realizados ensaios de genotoxicidade, ensaios *in vivo* e estudos farmacológicos de avaliação do perfil farmacocinético e farmacodinâmico (HUGHES *et al.*, 2011).

Só após a escolha dos candidatos a fármacos é que se prossegue para os ensaios clínicos, o último passo antes de os medicamentos atingirem o mercado. No entanto, esta barreira mostra-se difícil de transpor e uma grande parte dos projetos tende a falhar nesta etapa, mais precisamente nas fases II e III, onde é feita uma análise da eficácia terapêutica e da toxicidade do fármaco num determinado grupo de doentes (ÁLVAREZ-MACHANCOSES e FERNÁNDEZ-MARTÍNEZ, 2019).

Várias possibilidades são apontadas como motivo desta taxa de insucesso, entre as quais a falta de conhecimento adequado sobre os alvos e os mecanismos da doença, o que torna difícil o *design* de um fármaco e a previsão dos seus efeitos adversos (ÁLVAREZ-MACHANCOSES e FERNÁNDEZ-MARTÍNEZ, 2019), ou uma ineficaz avaliação das propriedades farmacocinéticas das moléculas.

Desta forma, é visível que o processo atual de descoberta e desenvolvimento de novos fármacos mostra-se pouco capaz de produzir os resultados esperados, surgindo assim a necessidade de encontrar novas estratégias que melhorem a sua eficácia, por exemplo na seleção de candidatos a fármacos mais promissores em fases mais prematuras do

desenvolvimento como forma de minimizar as falhas nas fases mais tardias. E é aqui que a IA se pode tornar uma mais valia.

### 3. Inteligência Artificial

---

Quando ouvimos falar em IA uma das primeiras ideias que nos surge é a de uma tecnologia futurista, de algo ligado à ficção científica. Na realidade, esta é uma tecnologia que já se encontra bem presente no nosso quotidiano e nas mais diversas áreas, como por exemplo no reconhecimento de imagens ou discursos e no marketing (SHAHID e LI, 2019). Esta tecnologia permitiu, até, a criação de algoritmos capazes de competir com o ser humano, como é o caso do programa Alpha Go que, após ter aprendido a jogar Go, um complexo jogo de estratégia, foi capaz de derrotar um jogador profissional e campeão europeu (SILVER *et al.*, 2016).

A IA pode ser definida como a demonstração de inteligência ou capacidade cognitiva por parte de máquinas e computadores, nomeadamente através da resolução de problemas ou da aprendizagem, características comumente associadas aos humanos (HESSLER e BARINGHAUS, 2018). Embora este seja um termo contemporâneo, a sua origem oficial remonta a 1956, durante um *workshop* decorrido na Universidade de Dartmouth e organizado pelo investigador John McCarthy e os seus colegas, onde o principal objetivo era tentar perceber qual a melhor maneira de construir máquinas capazes de simular a inteligência humana (AI100, 2016). Contudo, antes da oficialização do termo em si, já existiam investigadores que davam os primeiros passos e contribuía para este campo científico como foi o caso de Alan Turing que já em 1950 abordava a possibilidade de os computadores poderem aprender por si e referia a necessidade de aprender a programar estes mesmos computadores para que pudessem demonstrar inteligência (TURING, 1950).

Embora fossem inovadoras para a época, estas ideias enfrentaram à data algumas barreiras que impediram a sua passagem do campo teórico para o prático, entre as quais a falta de memória e de capacidade de processamento por parte dos computadores (AI100, 2016), levando a uma perda de interesse nesta tecnologia até ao surgimento da internet. A criação da internet, que permitiu a recolha de gigantescas quantidades de dados, e os avanços tecnológicos presentes nos dias de hoje permitiram que o interesse pela IA renascesse (AI100, 2016).

### 3.1 Machine Learning e Deep Learning

Existem dois conceitos importantes que fazem parte do universo da IA e que apesar de estarem relacionados entre si, são ao mesmo tempo diferentes. São esses conceitos o *Machine Learning* (ML) e o *Deep Learning* (DL).

Começando pelo ML, esta é uma área da IA que engloba um diverso número de algoritmos cujo principal objetivo passa pela elaboração de previsões a partir de um conjunto de dados (NICHOLS *et al.*, 2019) e pela procura de relações e padrões entre uma série de variáveis (BAŞTANLAR e ÖZUYSAL, 2014). Tal como a mente humana precisa de obter informações pertinentes para formular uma opinião adequada sobre um determinado assunto, também os algoritmos de ML, por vezes designados como modelos, necessitam de grandes quantidades de informação com o intuito de otimizar a sua capacidade de previsão (HESSLER e BARINGHAUS, 2018). Este processo de otimização do modelo, designado por *training*, consiste no fornecimento de consideráveis volumes de dados processados (*training data*) de forma a que o algoritmo aprenda com essa informação e faça previsões cada vez mais acertadas (BAŞTANLAR e ÖZUYSAL, 2014). Sendo assim, a primeira preocupação no *training* de um algoritmo passa por selecionar quais os *inputs* a incluir no mesmo. Estes *inputs*, comumente designados por características ou *features*, devem ser relevantes para o problema em questão e encontrar-se num formato que possa ser utilizado pelo modelo. Geralmente, esta triagem de dados deve ser feita com recurso à opinião de especialistas (BAŞTANLAR e ÖZUYSAL, 2014).

Apesar de serem necessárias grandes quantidades de dados para o *training* de um algoritmo de ML, é importante reservar uma fração independente desses dados como forma de testar o modelo no final, permitindo assim perceber qual a sua eficácia e identificar ajustes necessários à redução dos seus vieses (BAŞTANLAR e ÖZUYSAL, 2014).

Embora existam inúmeros modelos de ML com diferentes finalidades e estruturas, estes podem no geral ser incluídos em três grandes grupos: os modelos de *supervised learning*, de *unsupervised learning* e, por último, de *reinforcement learning*.

No primeiro grupo (*supervised learning* ou aprendizagem supervisionada), tal como o nome indica, há um controlo de processo de aprendizagem, ou seja, para além de se fornecerem os *inputs* ao algoritmo, também se estabelece quais as respostas que este tem de dar no final, algo que é bastante útil quando o objetivo é classificar algo. Por sua vez, no caso do *unsupervised learning* ou aprendizagem não supervisionada, apenas se fornecem os dados iniciais ao modelo, e cabe a este agrupar a informação e estabelecer padrões por si mesmo, agrupando resultados que, à partida, são similares (NICHOLS *et al.*, 2019). Por último, temos

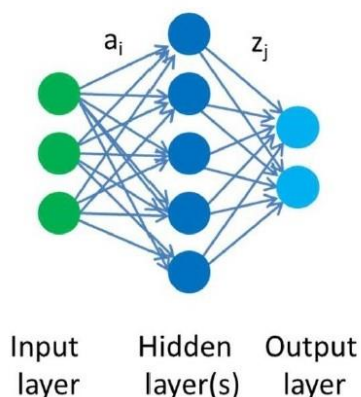
o grupo do *reinforcement learning*, também designado por aprendizagem por reforço, onde os modelos são treinados com base num sistema de recompensas. Com base no *outcome* ou resultado produzido pelo algoritmo, este recebe uma recompensa caso o resultado seja positivo, ou uma penalização, caso seja negativo, permitindo desta forma que o modelo melhore a sua capacidade de decisão e procure atingir recompensas cada vez mais altas (MAK e PICHKA, 2019).

O outro conceito a referir, e que constitui uma subárea do ML, é o DL, um conjunto de diferentes algoritmos que evoluíram a partir de um algoritmo muito estudado de ML, a *Artificial Neural Network* (ANN), também denominada de Rede Neuronal Artificial (CHEN et al., 2018). A ANN é um modelo que pretende mimetizar não só a estrutura como também a atividade do cérebro humano: o algoritmo possui inúmeras unidades designadas neurónios, que estão conectados entre si e dispostos em três tipos de camadas diferentes – a *input layer*, a *hidden layer* e a *output layer* – denominadas por camada de entrada, camada escondida e camada de saída respetivamente, tal como demonstrado na Fig. 1 (LAVECCHIA, 2015). As diferentes camadas são constituídas por diferentes números de unidades e cada neurónio recebe um sinal dos neurónios da camada anterior, sendo que esse sinal é multiplicado por um peso (*weight*) e posteriormente adicionado um valor designado de *bias* ou viés. O valor obtido é então processado por uma função de ativação e mediante o resultado, o neurónio é ou não ativado e transmite para a próxima unidade (LAVECCHIA, 2015). É importante referir que os valores dos pesos e dos vieses variam de neurónio para neurónio e é durante o treino do algoritmo que estes valores são ajustados de forma a fornecerem melhores resultados (HESSLER e BARINGHAUS, 2018).

Sendo assim, a grande diferença entre um modelo de DL, referido como *Deep Neural Network* (DNN) e as ANN é o nível de complexidade estrutural, visto que enquanto as últimas possuem apenas uma a duas *hidden layers*, que são as camadas onde as variáveis são transformadas até chegar à camada final, as DNNs possuem um maior número das mesmas (CHEN et al., 2018). Este aumento de complexidade permite não só fazer combinações mais sofisticadas como também aprender a partir de características mais abstratas (GAWEHN, et al., 2016), sendo que é aqui que se encontra a maior vantagem deste tipo de redes neuronais comparativamente às redes neuronais mais superficiais e outros modelos de ML, o facto de as DNNs conseguirem aprender a partir de dados não processados ou *raw data* (MAMOSHINA et al., 2016).

Embora seja uma tecnologia promissora, a DL ainda possui algumas limitações, nomeadamente o facto de exigir quantidades de dados superiores às do ML, a necessidade de

um alto poder computacional para um treino que é intensivo e com consumo de tempo elevado e o problema do “Black Box” (MAMOSHINA *et al.*, 2016), que será posteriormente abordado nesta monografia.



**Figura 1** Esquema da estrutura de uma rede neuronal: cada neurónio de uma camada está conectado a todos os neurónios da camada seguinte. Em regra, as ANNs são mais simples e possuem apenas até três *hidden layers*, no entanto, no caso das DNNs este número de camadas é muito superior, elevando assim a complexidade do algoritmo. (Adaptado de Hessler e Baringhaus, 2018).

### 3.2 Variedade de Algoritmos

A IA é uma vasta área tecnológica e, embora o objetivo de forma geral seja a prospeção de dados, os modelos que a constituem possuem várias estruturas e graus de complexidade diferentes, o que promove a possibilidade de múltiplas abordagens para um mesmo problema.

Os algoritmos de ML menos complexos possuem geralmente uma função de classificação, isto é, a partir da informação obtida procuram, ou atribuir uma categoria pré-estabelecida ou agruparem os compostos considerados similares em diferentes grupos. Alguns destes modelos são referidos ao longo desta monografia como é o caso do *support vector machine* (SVM) e do *k-nearest neighbor* (KNN).

Um dos mais simples é o KNN que faz a classificação com base na similaridade, agrupando moléculas com características similares num mesmo espaço, sendo que qualquer novo *input* é classificado consoante a proximidade com os pontos disponibilizados durante o treino (SHEN *et al.*, 2003). Por sua vez, o SVM é um modelo de *supervised learning* que funciona relativamente bem com pequenas quantidades de dados e que, através de métodos de regressão, é utilizado para classificação binária como, por exemplo, verificar se um composto é farmacologicamente ativo ou inativo (LAVECCHIA, 2015).

No entanto, ao entrar na subárea do DL, a complexidade dos modelos é muito superior. Como já foi referido anteriormente, este tipo de algoritmos evoluiu das ANN, sendo

constituídos por inúmeros neurónios interligados entre si e distribuídos por diferentes camadas. Para além da estrutura base das DNN onde cada neurónio está ligado aos neurónios da camada anterior e posterior, existem outras arquiteturas de modelos disponíveis, duas das quais referidas ao longo da monografia: a *recurrent neural network* (RNN) ou rede neuronal recorrente e o *autoencoder* ou autocodificador.

A RNN (Fig.2) é uma rede neuronal onde, em vez de a informação circular apenas de forma unidirecional de camada em camada até obtenção de um *outcome*, os neurónios transmitem informação a si mesmos, algo que não acontece nas outras estruturas de DL. Desta forma, ao contrário dos outros métodos onde a ordem dos dados analisados é irrelevante, o algoritmo consegue lidar com informação sequencial (CHEN *et al.*, 2018).

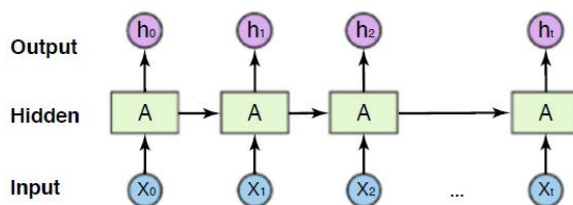
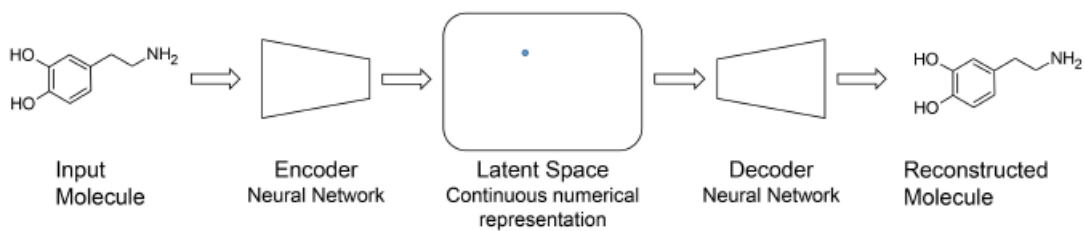


Figura 2 Estrutura de uma RNN. (Adaptado de Chen *et al.*, 2018).

Numa outra abordagem existem os *autoencoders* (Fig. 3). Este modelo é constituído por um codificador e um decodificador, sendo ambos formados por uma ou mais DNN. O codificador, como o nome indica, codifica a informação recebida numa forma mais compacta e por sua vez o decodificador procura restituir a informação original com base nos dados codificados (GÓMEZ-BOMBARELLI *et al.*, 2018). Desta forma, o objetivo passa por comprimir os dados numa forma latente de menores dimensões mantendo a informação pertinente. Este racional permite aplicar esta estratégia de forma adaptada em modelos geradores de novas moléculas onde, em vez de compactar um composto num único ponto do espaço latente, como um ponto caracterizado por coordenadas, o mesmo é codificado na forma de distribuição. Assim, durante a decodificação com recursos a métodos estatísticos, o decodificador procura restituir a molécula original possuindo ao mesmo tempo a capacidade de gerar estruturas similares com características diferentes, possibilitando uma melhor exploração do espaço latente e obtenção de potenciais novos compostos (BLASCHKE *et al.*, 2018). Esta estratégia adaptada dos *autoencoders* designa-se *variational autoencoder* (VAE).



**Figura 3** Estrutura de um *autoencoder*. O codificador tem como objetivo codificar uma molécula num ponto do espaço latente e em contrapartida o decodificador procura restituir a molécula original através da forma latente. (Adaptado de Blaschke *et al.*, 2018).

#### 4. Inteligência Artificial na Descoberta de Novos Fármacos

---

Um pilar fundamental para a aplicação da IA é a disponibilidade de grandes quantidades de informação, algo que se tem vindo a verificar nos últimos anos com quantidades colossais de dados sobre os sistemas biológicos a serem geradas, abrangendo diversas áreas nomeadamente a genómica ou a proteómica (VAMATHEVAN *et al.*, 2019). Em 2015, o Instituto Europeu de Bioinformática possuía 75 petabytes de informação biológica disponível para a investigação científica, sendo que esse valor tem tendência a aumentar de ano para ano (BROWN *et al.*, 2018), forçando assim a procura de métodos computacionais, ou *in silico*, que consigam de forma eficiente aproveitar todos os dados gerados, complementando assim os métodos experimentais tradicionais e reduzindo o tempo, o esforço humano e os custos dos mesmos (MACALINO *et al.*, 2018).

Apesar do considerável incremento de dados, o processo de descoberta e desenvolvimento de novos fármacos não conseguiu evoluir o suficiente de forma a tirar o máximo partido da quantidade de nova informação disponível, continuando a ser ineficaz em trazer novas moléculas para o mercado. Surge assim a oportunidade de aplicação da IA, na medida em que promove uma melhor análise de dados e complementa os métodos experimentais tradicionais.

Embora os modelos de IA possam ser aplicados a todas as fases da descoberta e desenvolvimento de fármacos (VAMATHEVAN *et al.*, 2019), nesta monografia serão abordadas algumas das possíveis aplicações nas fases iniciais do desenvolvimento de novas substâncias ativas, nomeadamente: a previsão das propriedades farmacocinéticas e toxicidade de compostos, a utilização em *design de novo* e o planeamento de síntese.



## 4.1 Previsão das Propriedades Farmacocinéticas e da Toxicidade de Compostos

Um potencial candidato a fármaco não só precisa de possuir afinidade para um alvo biológico como também precisa de reunir um leque de características que o tornem *drug-like* como por exemplo, boas características físico-químicas, boas propriedades de ADME (absorção, distribuição, metabolismo, excreção) e seletividade relativamente a alvos indesejados (HESSLER e BARINGHAUS, 2018).

Nos últimos anos, diversos modelos de ML foram desenvolvidos de forma a preverem as propriedades farmacocinéticas de compostos (DONIGER *et al.*, 2002; SHEN *et al.*, 2003). Um exemplo disso foi o trabalho de DONIGER *et al.* (2002) que desenvolveu um modelo de SVM cujo objetivo consistia em prever a capacidade de um composto conseguir atravessar a barreira hematoencefálica. Para o treino do modelo foram utilizados 324 compostos sendo cada molécula representada por 9 características (*features*) que englobavam propriedades físicas e estruturais, entre as quais o peso molecular e o número de ligações de hidrogénio. Após o treino, 30 conjuntos de dados de teste foram utilizados para validação e os *outputs* obtidos foram encorajadores, sendo possível classificar corretamente 82,7% dos compostos que atravessavam a barreira hematoencefálica e 80,2% dos compostos que não a conseguiam atravessar (DONIGER *et al.*, 2002).

Num outro estudo, SHEN *et al.* (2003) treinaram um modelo de KNN com a finalidade de avaliar a estabilidade metabólica de compostos, utilizando para a fase de treino um grupo de 631 compostos e os respetivos *turnovers* metabólicos hepáticos. Como possíveis resultados, o modelo classificava que compostos com menos de 25% de *turnover* eram considerados metabolicamente estáveis e acima de 75% metabolicamente instáveis, possuindo ainda duas classificações intermédias (ligeiramente estável e ligeiramente instável). No final, o modelo classificou corretamente 40 compostos como estáveis num total de 48 e 2 compostos como instáveis metabolicamente num conjunto de 6 compostos. Apesar da dificuldade em classificar moléculas com um *turnover* intermédio, o modelo mostrou ser bastante capaz de classificar as categorias extremas com um desempenho aceitável (SHEN *et al.*, 2003).

Na criação destes modelos de classificação surge, por vezes, uma dificuldade associada a conjuntos de dados desequilibrados, isto é, onde se possuem mais exemplos de uma determinada classe do que de outra (YANG *et al.*, 2018). Como forma de ultrapassar este desafio, YANG *et al.* (2018) criou um modelo designado *adaptive ensemble classification framework* (AECF), baseado num método *ensemble* ou de conjunto, onde o objetivo era prever não um único *outcome* mas sim 4 propriedades diferentes, nomeadamente: permeabilidade

através de células Caco-2, absorção intestinal humana, biodisponibilidade oral e ligação à glicoproteína-P (como substrato ou inibidor). De maneira a equilibrar os dados que possuíam, foi realizado um pré-tratamento que incluía técnicas como aumentar o número de compostos de classes menos representadas e reduzir o número de compostos de classes bem representadas. Depois, seguindo a estrutura normal de um método *ensemble*, os dados foram divididos e aplicados em diversos modelos individuais de ML como o KNN, o SVM ou *Random Forest*, sendo feita uma agregação dos valores obtidos em cada modelo individual. Por fim, o AECF obteve uma *performance* superior comparativamente a alguns dos métodos individuais utilizados na previsão das propriedades pretendidas.

Atualmente, com o desenvolvimento e criação dos mais diversos modelos, existem ferramentas *in silico* que passam a estar disponíveis em plataformas online como é o caso do ADMETlab (DONG *et al.*, 2018). O ADMETlab é uma ferramenta versátil que permite, entre outras modalidades, avaliar alguns parâmetros farmacocinéticos e de toxicidade de forma individual ou então fazendo uma avaliação sistemática de todas as propriedades para um determinado composto. DONG *et al.* (2018) recorreu a dados de 288961 compostos, a 9 métodos de regressão e 22 métodos de classificação para criar uma ferramenta que consegue realizar previsões para diferentes propriedades da ADME e toxicidade, estando as mesmas especificadas na Tabela 1.

**Tabela 1** Propriedades de ADME e Toxicidade avaliadas pela ferramenta ADMETlab. (Abreviaturas: CYP, Citocromo P450; DL50, Dose Letal 50; hERG, human-ether-a-go-go-related gene).

<b>Parâmetro</b>	<b>Propriedade(s)</b>
<b>Absorção</b>	Permeabilidade células Caco-2; Substrato Glicoproteína-P; Inibidor Glicoproteína-P; Biodisponibilidade
<b>Distribuição</b>	Ligação às proteínas plasmáticas; Permeabilidade da Barreira Hematoencefálica; Volume de Distribuição
<b>Metabolismo</b>	Inibidor ou substrato da CYP1A2; CYP3A4; CYP2C9; CYP2C19; CYP2D6
<b>Excreção</b>	<i>Clearance</i> ou depuração; Tempo de semi-vida
<b>Toxicidade</b>	Bloqueio dos receptores hERG; Teste de Ames; Hepatotoxicidade Humana; Sensibilização da pele; Dose Diária Máxima Recomendada, Lesão hepática provocada por fármacos e DL50

Esta ferramenta, tal como outros modelos criados com a mesma finalidade, permite, de forma rápida e menos dispendiosa, efetuar uma análise preliminar do perfil farmacocinético de determinados compostos no início do processo de desenvolvimento de fármacos, o que

permite não só excluir as moléculas que possuem perfis mais desfavoráveis como também ajuda a perceber características das mesmas que necessitam de otimização (DONG *et al.*, 2018).

Outra propriedade muito associada ao insucesso nos ensaios clínicos por motivos de segurança e que, portanto, é um importante parâmetro a ter em conta aquando do refinamento e seleção de compostos *lead*, é a potencial toxicidade dos mesmos (ÁLVAREZ-MACHANCOSES e FERNÁNDEZ-MARTÍNEZ, 2019).

Com os problemas éticos que vêm sendo levantados relativamente à utilização de animais em métodos *in vivo* e com as limitações associadas à utilização de métodos *in vitro*, como o tipo de células a utilizar ou a dificuldade de extrapolar os dados obtidos para os sistemas biológicos, têm surgido abordagens computacionais de IA como forma de avaliar de forma mais eficiente a toxicidade dos compostos (ZHANG *et al.*, 2018). Uma dessas abordagens, realizada por HUGHES e SWAMIDASS (2017), procurou antecipar e detetar a formação de um grupo de metabolitos por parte do citocromo P450 e peroxidases denominados por quinonas e que se ligam covalentemente a proteínas e DNA atuando como possíveis promotores de processos cancerígenos ou respostas imunológicas adversas. Utilizando um conjunto de 718 moléculas (formadoras de quinonas e não formadoras), foi treinada uma DNN com a capacidade não só de prever se a molécula tem, ou não, potencial para formar quinonas como também para identificar os possíveis locais na estrutura da mesma onde essa formação poderá ocorrer, mostrando a região da estrutura em causa com um padrão de cores ao qual está associado um *score* mediante a probabilidade de formação de quinonas (HUGHES e SWAMIDASS, 2017). Desta forma, este algoritmo possui o benefício de expor potenciais zonas de intervenção para uma possível otimização da molécula, não se focando apenas na avaliação da toxicidade.

O facto de os mecanismos de toxicidade serem muito complexos, influenciados por múltiplas variáveis e por vezes até nem serem conhecidos de todo, tem sido um desafio para a aplicação de algoritmos de IA com capacidade de prever o perfil de toxicidade das moléculas (MALTAROLLO *et al.*, 2014).

Os modelos de IA, quer para a previsão de propriedades de ADME quer para análise do perfil de toxicidade, não têm surgido como substitutos dos métodos experimentais, mas sim como um complemento dos mesmos e, embora requeiram melhorias adicionais e não consigam abranger todo o espaço químico ou avaliar todos os compostos, já conseguem fazer previsões suficientemente úteis para determinar, em fases mais iniciais do desenvolvimento de fármacos, quais as moléculas mais promissoras (GLEESON *et al.*, 2012).

## 4.2 Utilização em *Design De Novo*

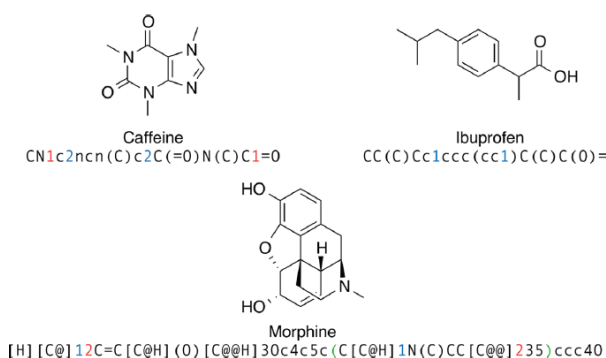
O número de compostos conhecidos atualmente, tendo em conta bases de dados privadas e de acesso público, ronda valores próximos de  $10^8$  compostos. No entanto, estima-se que, em teoria, seja possível existir um universo de  $10^{60}$  moléculas com potencial para serem consideradas *drug-like* (REYMOND *et al.*, 2012), o que constitui um grande desafio quando se fala em design de fármacos, uma vez que este vasto número de possíveis estruturas torna impraticável a realização de uma análise rigorosa de todos os elementos, mesmo recorrendo a *screening virtual*, isto é, recurso a abordagens computacionais com o intuito de analisar bibliotecas de compostos na procura de moléculas com capacidade de ligação a um determinado alvo. (SEGLER *et al.*, 2018).

No entanto, a evolução dos algoritmos de IA permitiu o desenvolvimento de uma outra área de aplicabilidade da mesma, o *design de novo*. Esta abordagem computacional tem por objetivo gerar novas entidades químicas que possuam um conjunto de características desejadas (MERK *et al.*, 2018) sendo que, em vez de uma pesquisa exaustiva e de certa forma aleatória do espaço químico, pretende-se incorporar o máximo de conhecimento sobre compostos com determinada ação biológica de forma a restringir a procura de novos compostos ativos a regiões de interesse desse mesmo espaço (HARTENFELLER e SCHNEIDER, 2011).

Uma interessante abordagem de um modelo gerador foi apresentada por MERK *et al.* (2018) que, com recurso a um algoritmo de RNN procurou obter compostos bioativos com atividade agonista para os recetores retinoides X e/ou recetores PPAR (*peroxisome proliferator-activated receptors*). A estratégia deste modelo consiste, numa primeira fase, em treinar o modelo com um conjunto inespecífico de compostos para que este aprenda a gerar moléculas válidas e, numa segunda fase, afinar o mesmo com recurso a um conjunto restrito de estruturas com atividade agonista para os recetores em causa, de forma a serem geradas estruturas similares às de referência. Para este treino ser possível, os compostos precisam de ser representados num formato que possa ser interpretado pelos modelos, sendo, na maioria dos casos, o formato utilizado o *Simplified Molecular Input Line Entry System* ou SMILES (Fig. 4). Este formato consiste em representar uma molécula numa única linha recorrendo a letras, números e símbolos, simplificando o *input* de informação (SEGLER *et al.*, 2018). O treino foi executado com recurso a 54155 SMILES de compostos bioativos e a afinação do algoritmo foi realizada com recurso a 25 moléculas de ácidos gordos contendo um grupo carboxilo e com atividade agonista para os recetores em estudo. A partir deste modelo, obtiveram-se 1000 SMILES, sendo que 93% correspondiam a moléculas válidas e 90% representavam estruturas únicas sem duplicados. Com base nestes resultados, foi executada uma classificação via computacional

dos mesmos de forma a prever a capacidade de modular os recetores retinoides X e/ou os recetores PPAR, obtendo-se 49 moléculas com um score favorável. Foram então seleccionados, tendo em conta a classificação *in silico* e a disponibilidade de blocos de construção (*building blocks*), 5 compostos para síntese, de forma a serem realizados ensaios *in vitro*. No final, verificou-se que 2 desses compostos eram agonistas para ambos os tipos de recetores e que outros 2 apenas eram agonistas dos recetores PPAR.

Tal como pode ser observado, o sucesso deste tipo de estratégia funcionar parece depender da existência prévia de compostos ativos para determinado alvo. Neste contexto e utilizando também um algoritmo de RNN, SEGLER *et al.* (2018) verificou inicialmente que, após utilizar compostos ativos conhecidos para a afinação, um modelo gerador consegue gerar moléculas análogas ou similares às que serviram de referência. No entanto, SEGLER *et al.* (2018) procurou desafiar o mesmo modelo na tentativa de criar compostos ativos sem a utilização de moléculas conhecidas para determinado alvo. Desta forma, treinaram o modelo com recurso a 1,4 milhões de estruturas SMILES e, após essa fase, geraram cerca de 100 mil moléculas. De seguida, recorrendo a um algoritmo de ML, previram se os compostos obtidos eram ativos, tendo como alvo a bactéria *Staphylococcus aureus*, e utilizaram as estruturas consideradas ativas para uma afinação do algoritmo de RNN. No final, reunindo as moléculas obtidas antes e após a afinação, o modelo conseguiu gerar perto de 61 mil moléculas com potencial atividade para o alvo em causa. O que demonstra que, mesmo não possuindo compostos ativos conhecidos *a priori*, a junção de um modelo gerador com um modelo de previsão pode ser uma estratégia vantajosa. Contudo, em *design de novo*, esta é apenas uma das estratégias possíveis de serem utilizadas.



**Figura 4** Representação SMILES das moléculas de Cafeína, Morfina e Ibuprofeno. (Adaptado de Segler *et al.* 2018).

Uma outra abordagem mais recente e com grande potencial é o recurso a modelos VAE. GÓMEZ-BOMBARELLI *et al.* (2018) aplica este conceito na tentativa de criar ferramentas de exploração do espaço latente, consoante a propriedade que se pretende

estudar. Com esse objetivo, treinaram um VAE com o intuito de criar um espaço latente a partir do qual poderiam gerar moléculas de um ponto aleatório, acrescentando ao mesmo um modelo capaz de prever as propriedades do composto tendo em conta a sua representação nesse espaço. Assim, o espaço latente é reorganizado de acordo com a propriedade em estudo onde, a partir de um determinado ponto, podemos explorar este espaço na direção que melhor favorece o objetivo, descodificando depois as representações latentes em estruturas moleculares. No entanto, para que um modelo generativo com base num espaço latente funcione, é preciso que este mantenha o princípio da similaridade, isto é, moléculas com estrutura similar devem encontrar-se próximas nesse mesmo espaço (BLASCHKE *et al.*, 2018). Por esse motivo, GÓMEZ-BOMBARELLI *et al.* (2018) utilizou a representação do ibuprofeno no espaço latente como orientação e verificou que, quanto mais distante no espaço as moléculas estavam, menos similares eram ao ibuprofeno, o que demonstra que o princípio se manteve. De seguida, recorreu a um modelo de previsão de propriedades para reestruturar o espaço latente, de modo a que moléculas com maior potencialidade se localizem numa determinada região e moléculas de características menos desejadas noutra. Com recurso a um processo gaussiano (modelo de previsão selecionado), estabeleceram como objetivo da otimização a procura da molécula mais *drug-like* e com melhor possibilidade de síntese. Esta é uma estratégia muito mais vantajosa não só porque origina moléculas em percentis superiores a outros métodos de procura, mas também porque o processo de otimização a partir de uma determinada molécula faz com que a molécula final esteja localizada numa região onde o valor do objetivo de estudo é superior. Embora esta estratégia ainda possua algumas limitações como por exemplo algumas das SMILES geradas não serem válidas, esta associação entre um VAE e um modelo preditivo mostrou ser uma ferramenta de grande potencial na exploração de um espaço latente na procura de compostos.

Com uma abordagem semelhante à apresentada anteriormente, ZHAVORONKOV *et al.* (2018) criou um modelo designado *Generative Tensorial Reinforcement Learning (GENTRL)* com o objetivo de encontrar inibidores dos recetores DDR1, um recetor pró-inflamatório associado a processos de fibrose. Desta forma, o modelo baseia-se na utilização de um VAE em conjunto com técnicas de *reinforcement learning*. Desta forma, organizou inicialmente um espaço latente com recurso a um vasto conjunto de moléculas, incluindo também, mas não só, inibidores conhecidos dos recetores DDR1 e inibidores comuns de outras cinases. Depois, utilizou o *reinforcement learning* com recurso a 3 algoritmos que atuam como filtros. Um classificou as moléculas geradas quanto à sua novidade (com base em moléculas já patenteadas), outro fez a distinção entre inibidores de cinases e outros tipos de moléculas e

por fim o último fez a triagem dos inibidores específicos dos recetores DDR1. No final, 40 moléculas foram selecionadas, onde 6 dessas moléculas foram selecionadas para validação do modelo. Após realização de ensaios *in vitro* para avaliar a atividade inibitória dos compostos, dois compostos apresentaram uma forte capacidade inibitória, possuindo uma concentração inibitória para metade do efeito máximo de 10 e 21 nM. Estudos posteriores a esses mesmos compostos verificaram que não só possuíam propriedades físico-químicas satisfatórias como também apresentavam uma *clearance* e tempo de semivida semelhantes a outras moléculas usadas para o mesmo alvo. Embora estes compostos testados possuam potencial para serem candidatos a fármacos, eles poderão requerer ainda otimização quer a nível de seletividade, especificidade ou outras propriedades, não obstante o facto de todo o processo de descoberta, síntese e validação das moléculas para os recetores DDR1 ter sido desenvolvido num período inferior a 2 meses (46 dias), demonstra a potencialidade dos modelos generativos com base em algoritmos de DL no design e desenvolvimento de novos fármacos.

É importante entender que é improvável os compostos gerados por técnicas de design *de novo* reunirem, à primeira tentativa, as condições ideais tanto a nível de atividade, especificidade, seletividade e perfil farmacocinético, mas que estes compostos, com uma otimização apropriada *a posteriori*, constituem um ponto de partida para se obter um maior número de hits, promovendo a rapidez e eficiência do processo de desenvolvimento de fármacos (HARTENFELLER e SCHNEIDER, 2011).

### **4.3 Planeamento de Síntese**

Ao longo desta monografia, foi sendo demonstrado que os métodos de IA podem ser úteis não só na previsão das propriedades farmacocinéticas e da toxicidade de compostos como também no design de novas moléculas. Contudo, num processo de descoberta e desenvolvimento de fármacos, por muito que um composto seja promissor, este só poderá ser realmente relevante se for sinteticamente acessível, de modo a que uma análise experimental possa ser realizada.

O crescente interesse nos algoritmos de IA nos últimos anos aliado a uma maior disponibilidade de informação relativamente a inúmeras reações químicas permitiu o aparecimento de modelos cujo objetivo consiste em prever como determinados *building blocks* vão reagir ou em identificar quais as vias de síntese mais adequadas para se obter um produto de reação pretendido (ENGKVIST *et al.*, 2018).

Dentro do planeamento de síntese, existem dois principais componentes que, embora diferentes, se encontram intimamente relacionados entre si: a previsão de reações e a retró síntese. A previsão de reações, tal como o nome indica, consiste em prever como é que

um conjunto de reagentes conhecidos vão interagir e que produto irão originar (SEGLER e WALLER, 2017). Por sua vez, a retróssíntese baseia-se no processo inverso onde, a partir de uma determinada molécula, se pretende descobrir quais os reagentes que provavelmente estarão envolvidos na sua síntese. Este componente é mais complexo visto que a informação com que se começa inicialmente é menor, uma vez que apenas se sabe o produto da reação, e há um grande número de possíveis sínteses (MATER e COOTE, 2019).

Os métodos tradicionalmente utilizados, tanto na análise retróssintética como na previsão de reações, baseiam-se em *ruled-based expert systems* (RBES), também designados como sistemas baseados em regras. Estes sistemas usam como fonte de informação um conjunto de regras ou *templates*, codificadas por especialistas ou oriundas de bases de dados, que determinam como é que os grupos funcionais estão envolvidos numa determinada reação. Desta forma, ao fornecer ao sistema uma molécula, este aplica as regras de reação na mesma como forma de originar, dependendo do caso, um produto ou reagentes (SEGLER e WALLER, 2017). Apesar de serem comuns, estes tipos de abordagens possuem várias limitações uma vez que conseguem fazer previsões de reações que não estejam contidas no seu conjunto de regras e ignoram o contexto geral da estrutura molecular focando-se apenas nas regiões do centro de reação, ignorando partes da molécula que possam interferir na reação (LIU *et al.*, 2017).

Fazendo referência inicialmente aos métodos de previsão de reações, a maioria das abordagens recorrem a uma estratégia híbrida onde unem a IA e as regras de reação utilizadas nos RBES. Um exemplo deste tipo de abordagem é demonstrado por SEGLER e WALLER (2017) que procuraram, recorrendo a uma rede neuronal de ML, prever qual seria a regra mais provável e apropriada para um determinado *input*, permitindo assim priorizar as regras a utilizar em vez de as testar uma a uma na totalidade. Numa fase inicial treinaram a rede neuronal com recursos a milhões de exemplos de reações e respetivas regras de modo a que o algoritmo conseguisse generalizar para outras moléculas, aprendendo a estabelecer padrões entre as estruturas moleculares dos compostos e as reações. Como forma de avaliar o desempenho do modelo híbrido, este foi comparado com um modelo de RBES normal que não recorria a IA. Numa primeira experimentação, os modelos tinham de prever qual a regra correta de um conjunto de 103 regras codificadas à mão. O modelo híbrido demonstrou uma precisão de 92%, um valor muito superior comparativamente à precisão de 7% do modelo RBES. O mesmo tipo de teste foi feito recorrendo, desta vez, a 8720 regras automaticamente extraídas. Nesta situação, o modelo híbrido apresentou uma precisão de 77% face a 7% do modelo RBES. Estes valores demonstram que o modelo híbrido tem uma *performance* muito



superior à dos modelos de RBES uma vez que dá prioridade às regras mais relevantes. No entanto, esta abordagem também herda algumas limitações dos modelos tradicionais como a dificuldade de prever reações que não se encontram na sua base de conhecimento ou não considerarem estereoisómeros.

Um outro exemplo de um modelo híbrido foi proposto por COLEY *et al.* (2017). Neste caso em particular, o objetivo passou por prever qual o produto principal de uma reação. Numa primeira fase, são utilizados cerca de 15 mil exemplos de reação, sendo que os reagentes dos mesmos constituem o conjunto de reagentes usados pelo modelo. De seguida, aplicam 1689 *templates* de reação ao grupo de reagentes e geram-se inúmeros candidatos a produtos de reação. Esta primeira fase permite gerar vários produtos possíveis para o mesmo conjunto de reagentes, algo que permite contornar uma limitação comum na quantidade de dados de reação disponíveis, visto que geralmente existe apenas informação sobre reações com rendimentos elevados. Na segunda fase, com recurso a uma rede neuronal de ML, o modelo procura prever qual é o produto principal da reação. No final, este modelo híbrido apresentou uma precisão de 71,8% em identificar o produto principal, isto é, o produto que está registado nas bases de dados como o verdadeiro produto de reação entre dois reagentes. Embora seja uma boa estratégia, este modelo apresenta também algumas limitações como, por exemplo, não ter em conta condições de reação que podem alterar o resultado da reação.

Relativamente à análise retrossintética, grande parte das abordagens são semelhantes às apresentadas anteriormente. O mesmo modelo apresentado por SEGLER e WALLER (2017) foi também aplicado a casos de retrossíntese e, embora tenha também demonstrado uma precisão de 78% face a 5% do modelo RBES, este modelo não é geralmente tão eficaz para retrossíntese uma vez que a informação de *input* é mais reduzida, sabendo-se inicialmente apenas qual o produto da reação.

Embora o paradigma atual seja baseado nesta fusão de técnicas, o facto de a maioria dos modelos não conseguirem propor desconexões inovadoras para uma molécula nem fazer previsões que não estejam abrangidas pela sua base de conhecimento, requer que se procurem novas abordagens.

Uma dessas abordagens é apresentada por LIU *et al.* (2017) que, excluindo qualquer utilização de RBES, procura utilizar uma abordagem de DL para encontrar os materiais de início mais prováveis para uma determinada molécula. Utilizando o formato SMILES para representar todas as moléculas, LIU *et al.* (2017) começou por reunir aproximadamente 50 mil exemplos de reação, onde cada exemplo correspondia a um único produto e que podia ser agrupado em 10 tipos diferentes de reação, incluindo oxidação, redução, adição de grupo

funcional, proteção, entre outros. Durante o treino, o *input* é constituído pela SMILES do produto de reação mais um indicativo do tipo de reação que irá ocorrer, e pretende obter-se as SMILES representativas dos reagentes de reação. Como forma de testar o modelo, este foi comparado com um modelo RBES que recorre a regras de retró síntese. No final, o algoritmo de LIU *et al.* (2017) apresentou um desempenho equiparável aos métodos tradicionais, onde, considerando as 50 melhores previsões de cada modelo, apresentou uma precisão de 70,7% face a 69,5% do modelo RBES. No entanto, esta performance varia conforme a quantidade de exemplos fornecidos e o tipo de reação em análise, sendo que o modelo proposto possuiu um melhor desempenho em reações de proteção e desproteção uma vez que consegue capturar de forma mais eficaz o ambiente molecular global das moléculas intervenientes na reação, algo que os modelos tradicionais não fazem, visto que se focam apenas no centro de reação. Esta é uma abordagem inovadora e que possui vantagens relativamente a outras abordagens híbridas visto poder ser treinada diretamente com dados de reações e considerar todos os constituintes das moléculas. Por outro lado, apresenta ainda algumas limitações como o facto de originar SMILES inválidas ou de propor reações que não são exequíveis.

Embora existam diversas estratégias para o planeamento da síntese, algumas das quais bastante promissoras e que podem influenciar a descoberta de novos fármacos, é necessária ainda uma evolução dos algoritmos de IA utilizados, sendo para isso fundamental que exista mais informação disponível nas bases de dados, não só sobre reações com altos rendimentos mas também sobre exemplos negativos, fundamentais para a aprendizagem dos algoritmos (MATER e COOTE, 2017).

## **5. Vantagens do Uso de Inteligência Artificial**

---

O paradigma atual da descoberta e desenvolvimento de novos fármacos pode ser descrito por dois adjetivos: moroso e dispendioso. Em adição, há uma tendência para que esta tendência evolua negativamente. Em 2001, os custos associados ao desenvolvimento de um medicamento rondavam os 800 milhões de dólares, no entanto, atualmente esses valores escalam para valores na ordem dos 3 mil milhões de dólares (MAK e PICHKA, 2019) sem que, em contrapartida, se verifique um aumento do número de novos fármacos aprovados (DONG *et al.*, 2018). A recente ascensão da IA e respetivos algoritmos surge como uma oportunidade de reverter esta situação.

A primeira grande potencialidade desta tecnologia, e a que abre as portas para as restantes vantagens da sua utilização, é a capacidade, e até necessidade, que os modelos de IA têm de processar grandes quantidades de dados. Nos dias de hoje existem quantidades colossais de dados a serem gerados continuamente, o que se torna um desafio para os investigadores que não conseguem assimilar toda esta informação sobre os sistemas biológicos e tirar o melhor proveito da mesma (BROWN *et al.*, 2018). Desta forma, estratégias de ML e DL permitem um melhor processamento e uma análise mais aprofundada dos dados existentes, facilitando o trabalho dos investigadores e permitindo até novas descobertas.

Graças a esta prospeção de dados, o uso de IA permite que as fases iniciais do processo de desenvolvimento ocorram mais rapidamente. Isto é visível, por exemplo, quando os algoritmos preveem as propriedades farmacocinéticas e a toxicidade das moléculas, selecionando apenas as mais promissoras, evitando assim um gasto adicional de tempo em moléculas que não reúnem os requisitos adequados (DONG *et al.*, 2018). Um outro exemplo possível de ser utilizado é o do *design de novo* onde, em vez de se realizar um *screening* arbitrário de bibliotecas de compostos, há um *design* de novas moléculas com potencial ação nos alvos desejados, facilitando a descoberta de novos compostos *hit*.

O recurso a IA permite também um aumento da segurança e eficácia das moléculas em estudo. Ao existir uma priorização dos compostos com melhores propriedades e menor toxicidade, o conjunto de estruturas que atingem os ensaios clínicos serão, consequentemente, mais seguros e com um perfil farmacocinético mais adequado, aumentando assim a probabilidade de sucesso (ÁLVAREZ-MACHANCOSES e FERNÁNDEZ-MARTÍNEZ, 2019).

Como consequência das anteriores surge um outro benefício: a redução dos custos. Embora seja necessário um investimento inicial nos recursos computacionais, esta tecnologia permite a redução de ensaios experimentais dispendiosos, diminuindo a quantidade de substâncias a sintetizar e assegurando que os compostos que atinjam os ensaios clínicos sejam os mais promissores, permitindo uma menor taxa de insucesso e consequentemente uma redução dos prejuízos associados ao insucesso (MAK e PICHKA, 2019).

Apesar de ser uma ferramenta promissora e que pode trazer muitos benefícios para a indústria farmacêutica, ainda existem barreiras que precisam ser ultrapassadas de forma a que a IA possa verdadeiramente revolucionar o processo de desenvolvimento de novos fármacos.

## 6. Desafios na Aplicação da Inteligência Artificial

---

Uma das principais dificuldades existentes na aplicação da IA foca-se em algo essencial ao sucesso desta tecnologia: dados. Independentemente do quão bem desenvolvido é um algoritmo, este só irá produzir resultados satisfatórios se as informações utilizadas como *input* forem relevantes (ESAKI *et al.*, 2019).

A primeira questão está relacionada com a qualidade dos próprios dados, uma vez que, na sua grande maioria, são oriundos de ensaios *in vitro* ou ensaios *in vivo* em animais e que, para além de nem sempre retratarem de forma adequada as condições fisiológicas normais do corpo humano, a forma como são obtidos é também variável de ensaio para ensaio. Sendo assim, há o risco de que algoritmos treinados com base nestes *inputs* nem sempre produzam resultados relevantes (KNUDSEN *et al.*, 2015). A segunda questão é relativa à quantidade de informações disponíveis. A IA necessita de um volume elevado de dados, incluindo não só exemplos positivos como também exemplos negativos. No entanto, verifica-se que os dados disponíveis atualmente são muito desequilibrados, isto é, existe mais informação sobre compostos promissores em detrimento de compostos menos ativos ou até ineficazes, resultando numa falta de informação sobre inúmeros compostos (LECUN *et al.*, 2015).

Atualmente, já existem bases de dados de acesso público que procuram juntar o máximo de informação sobre ensaios biológicos e resultados de fases pré-clínicas e clínicas do desenvolvimento de fármacos como forma de aumentar os *inputs* disponíveis para aplicar nos modelos de IA (GAULTON *et al.*, 2017), mas, apesar deste esforço, as bases de dados públicas representam apenas uma pequena fração da informação existente, sendo que a maioria constitui propriedade privada das empresas farmacêuticas. É, portanto, importante que exista uma maior partilha de informação para que se possa impulsionar com sucesso os algoritmos de ML e DL.

Um outro desafio relativo à implementação da IA está ligado à sua interpretabilidade. Existem modelos de ML e DL, em especial os últimos devido à sua complexidade estrutural, que são considerados “*black boxes*”, isto é, autênticas “caixas negras” nas quais entram grandes conjuntos de dados e de onde sai um resultado sem que por vezes se saiba qual o racional que originou essa conclusão (MAMOSHINA *et al.*, 2016). Para além disso, ao ser exposto a nova informação, o algoritmo pode fornecer um resultado diferente do original, o que pode provocar inúmeros prejuízos no processo de desenvolvimento de fármacos (PRICE, 2018). Desta forma, a falta de uma explicação científica para um determinado resultado e o facto desse mesmo resultado poder ser moldado mediante os *inputs* que são fornecidos ao sistema

constituem uma barreira na aceitação por parte dos investigadores e indústrias (SHAW *et al.*, 2019).

Por fim, existem desafios que não envolvem propriamente assuntos técnicos da IA mas que de igual forma dificultam a sua implementação. Um desses obstáculos é o investimento económico inicial necessário à obtenção e manutenção de recursos computacionais (SHAW *et al.*, 2019). O outro consiste na necessidade de recursos humanos adequados e adaptados para esta realidade. É necessário recrutar especialistas em IA e ciência de dados, algo difícil devido à escassez destes profissionais, e é preciso apostar na formação dos intervenientes no desenvolvimento de novos fármacos, para que estes consigam compreender esta tecnologia e permitir uma melhor simbiose da mesma com os métodos tradicionais (LANSDOWNNE, 2020). Neste contexto, acresce referir que a variedade e complexidade dos algoritmos de ML e DL pode por vezes constituir uma barreira à sua adoção.

## 7. Conclusão

---

Apesar de esta ser uma era de elevados avanços tecnológicos, o paradigma atual do desenvolvimento de fármacos parece não acompanhar esta tendência, continuando a ser um caminho cada vez mais longo e dispendioso, sem que se verifiquem melhores resultados. É, portanto, perceptível que são necessárias estratégias que aumentem a eficácia deste processo e que impulsionem esta área para um nível digno do século XXI.

A ascensão da IA nos últimos anos, promovida pelo desenvolvimento dos meios computacionais e pelo incremento da informação disponível, surge como um possível aliado no desenvolvimento de novos fármacos, sendo a sua aplicação exequível em todas as fases do processo. Nesta monografia, o foco de aplicação foi centrado em exemplos de três aplicações distintas mas igualmente importantes: a previsão de propriedades farmacocinéticas e toxicológicas de compostos como forma de identificar de forma mais rápida e eficaz moléculas promissoras para um determinado alvo; o *design* de novas entidades químicas através de uma melhor exploração do espaço químico, designado *design de novo*; e o planeamento de síntese de compostos, como forma de procurar vias viáveis para a obtenção de compostos que serão posteriormente testados. Em todas as situações analisadas, a aplicação desta tecnologia emergente não é vista como um substituto quer dos métodos tradicionais quer dos profissionais envolvidos, mas apenas como mais uma ferramenta, ainda dependente de intervenção humana, que procura aprimorar os processos atualmente utilizados, tornando-os mais eficazes, baratos e rápidos.

No entanto, a implementação desta nova ferramenta não está livre de desafios e cabe à indústria farmacêutica encontrar estratégias que os ultrapassem e tornem a IA presença habitual nos seus processos. O aumento da partilha de dados, tanto de acesso público como privado, como forma de aumentar a qualidade e quantidade dos dados disponíveis, o estudo dos algoritmos com vista a torná-los mais “transparentes” ou a criação de regulamentação que permita criar modelos que, embora não permitam perceber qual o seu processo de decisão, produzam resultados que possam ser considerados válidos e aceitáveis são algumas das estratégias que podem ser aplicadas para facilitar a adoção dos algoritmos de ML e DL.

Embora o potencial da IA seja claramente visível, os algoritmos existentes são ainda algo imaturos e isso reflete-se no facto de ainda não existir um único medicamento desenvolvido com base em abordagens que envolvam esta tecnologia (MAK e PICHKA, 2019). No entanto, o crescente interesse e investimento nesta ferramenta irá certamente acelerar a descoberta e desenvolvimento de novos fármacos, provocando alterações permanentes na indústria farmacêutica.

## 8. Referências Bibliográficas

---

ÁLVAREZ-MACHANCOSES, O., FERNÁNDEZ-MARTÍNEZ, J.L., - Using artificial intelligence methods to speed up drug discovery. *Expert Opinion on Drug Discovery*. 14, 8 (2019) 769-777.

BARNASH, K. D., JAMES, L. I., FRYE, S. V. – Target class drug discovery. *Nature Chemical Biology*. 13, 10 (2017) 1053-1056.

BAŞTANLAR, Y., ÖZUYSAL, M., - Introduction to Machine Learning. In: YOUSEF, M., ALLMER, J., *miRNomics: MicroRNA Biology and Computational Analysis*, New York: Humana Press, 2014, ISBN: 978-1-62703-748-8. p.105-128.

BATOOL, M., AHMAD, B., SANGDUN, C. – A Structure-based drug discovery paradigm. *International Journal of Molecular Sciences*. 20, 11 (2019).

BLASCHKE, T., OLIVECRONA, M., ENGVIST, O., BAJORATH, J., CHEN, H. – Application of generative autoencoder in de novo molecular design. *Molecular Informatics*. 37, 1 (2018) 1-11.

BLEICHER, K. H., BÖHM, H. J., MÜLLER, K., ALANINE, A. I. – Hit and lead generation: beyond high-throughput screening. *Nature Reviews Drug Discovery*. 2, 5 (2003) 369-378.

BROWN, N., CAMBRUZZI, J., COX, P. J., DAVIES, M., DUNBAR, J., PLUMBIEY, D., SELLWOOD, M. A., SIM, A., WILLIAMS-JONES, B. I., ZWIERZYNA, M., SHEPPARD, D. W. – Big data in drug discovery. *Progress in Medicinal Chemistry*. 57, 1 (2018) 277-356.

CHEN, H., ENGVIST, O., WANG, Y., OLIVECRONA, M., BLASCHKE, T. – The rise of deep learning in drug discovery. *Drug Discovery Today*. 23, 6 (2018) 1241- 1250.

COLEY, C. W., BARZILAY, R., JAAKKOLA, T. S., GREEN, W. H., JENSEN, K. F. – Prediction of Organic Reaction outcomes using machine learning. *ACS Central Science*. 3, 5 (2017) 434-443.

DONG, J., WANG, N., YAO, Z., ZHANG, L., CHENG, Y., OUYANG, D., LU, A., CAO, D. – ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *Journal of Cheminformatics*. 10, 1 (2018) 1-11.

DONIGER, S., HOFMANN, T., YEH, J. – Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms. *Journal of Computational Biology*. 9, 6 (2002) 849-864.

DREWS, J. – Drug discovery: a historical perspective. *Science*. 287, 5460 (2000) 1960-1964.

EUROPEAN FEDERATION OF PHARMACEUTICAL INDUSTRIES AND ASSOCIATION – The pharmaceutical industry in figures. 2019. [Consultado a 01 de junho de 2020]. Disponível na Internet: <https://www.efpia.eu/media/413006/the-pharmaceutical-industry-in-figures.pdf>

ENGVIST, O., NORRBY, P., SELMI, N., LAM, Y., PENG, Z., SHERER, E. C., AMBERG, W., ERHARD, T., SMYTH, L. A. – Computational prediction of chemical reactions: current status and Outlook. *Drug Discovery Today*. 23, 6 (2018) 1203-1218.

ESAKI, T., WATANABE, R., KAWASHIMA, H., OHASHI, R., NATSUME-KITATANI, Y., NAGAO, C., MIZUGUCHI, K. – Data curation can improve the prediction accuracy of metabolic intrinsic clearance. *Molecular Informatics*. 38, 1 (2019) 1-10.

GAULTON, A., HERSEY, A., NOWOTKA, M., BENTO, A. P., CHAMBERS, J., MENDEZ, D., MUTOWO, P., ATKINSON, F., BELLIS, L. J., CIBRIÁN-UHALTE, E., DAVIES, M., DEDMAN, N., KARLSSON, A., MAGARIÑOS, P., OVERINGTON, J. P., PAPADATOS, G., SMIT, I., LEACH, A. – The ChEMBL database in 2017. *Nucleic Acids Research*. 45, 1 (2017) 945- 954.

GAWEHN, E., HISS, J. A., SCHNEIDER, G. – Deep learning in drug discovery. *Molecular Informatics*. 35, 1 (2016) 3-14.

GLEESON, M. P., MODI, S., BENDER, A., ROBINSON, R. L. M., KIRCHMAIR, J., PROMKATKAEW, M., HANNONGBUA, S., GLEN, R. C. – The challenges involved in modeling toxicity data in silico: a review. *Current Drug Metabolism*. 18, 9 (2012) 1266-1291.

GÓMEZ-BOMBARELLI, R., WEI, J. N., DUVENAUD, D., HERNÁNDEZ-LOBATO, J. M., SÁNCHEZ-LENGELING, B., SHEBERLA, D., AGUILERA-IPARRAGUIRRE, J., HIRZEL, T. D., ADAMS, R. P., ASPURU-GUZIK, A. – Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*. 4, 2 (2018) 268-276.

HARTHENFELLER, M., SCHNEIDER, G. – De novo drug design. In: BAJORATH, J., *Chemoinformatics and Computational Chemical Biology*, Humana Press, 2011, ISBN: 978-1-60761-838-6, p.299- 323.

HESSLER, G., BARINGHAUS, K. H. – Artificial intelligence in drug design. *Molecules*. 23, 10 (2018)



HUGHES, J. P., REES, S., KALINDJIAN, S. B., PHILPOTT, K. L. – Principles of early drug discovery. *British Journal of Pharmacology*. 162, 6 (2011) 1239-1249.

HUGHES, T. B., SWAMIDASS, S. J. – Deep learning to predict the formation of quinone species in drug metabolism. *Chemical Research in Toxicology*. 30, 2 (2017) 642-656.

KNUDSEN, T. B., KELLER, D. A., SANDER, M., CARNEY, E. W., DOERRER, N. G., EATON, D. L., FITZPATRICK, S. C., HASTINGS, K. L., MENDRICK, D. L., TICE, R. R., WATKINS, P. B., WHELAN, M. – FutureTox II: in vitro data and in silico models for predictive toxicology. *Toxicological Sciences*. 143, 2 (2015) 256-267.

LANSDOWNE, L. E. – AI in pharma – adoption, disruption and drug discovery. 2020. [Consultado a 15 de maio de 2020] Disponível na internet: <https://www.technologynetworks.com/drug-discovery/articles/ai-in-pharma-adoption-disruption-and-drug-discovery-330256>

LAVECCHIA, A. – Machine-learning approaches in drug discovery: methods and applications. *Drug Discovery Today*. 20, 3 (2015) 318-331.

LECUN, Y., BENGIO, Y., HINTON, G. – Deep Learning. *Nature*. 521,7553 (2015) 436-444.

LIU, B., RAMSUNDAR, B., KAWTHEKAR, P., SHI, J., GOMES, J., NGUYEN, Q. L., HO, S., SLOANE, J., WENDER, P., PANDE, V. – Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Central Science*. 3, 10 (2017) 1103-1113.

MACALINO, S., BASITH, S., CLAVIO, N., CHANG, H., KANG, S., CHOI, S. – Evolution of in silico strategies for protein-protein interaction drug discovery. *Molecules*. 23, 8 (2018).

MAK, K., PICHIKA, M. R. – Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today*. 24, 3 (2019) 773-780.

MALTAROLLO, V. G., GERTRUDES, J. C., OLIVEIRA, P. R., HONORIO, K. M. – Applying machine learning techniques for ADME-Tox prediction: a review. *Expert Opinion on Drug Metabolism and Toxicology*. 11, 2 (2014) 259-271.

MAMOSHINA, P., VIEIRA, A., PUTIN, E., ZHAVORONKOV, A. – Applications of deep learning in biomedicine. *Molecular Pharmaceutics*. 13, 5 (2016) 1445-1454.

MATER, A. C., COOTE, M. L. – Deep learning in chemistry. *Journal of Chemical Information and Modeling*. 59 (2019) 2545-2559.

MERK, D., FRIEDRICH, L., GRISONI, F., SCHNEIDER, G. – De novo design of bioactive small molecules by artificial intelligence. *Molecular Informatics*. 37,1 (2018) 3-6.

NICHOLS, J. A., CHAN, H. W. H., BAKER, M. A. B. – Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophysical Reviews*. 11,1 (2019) 111-118.

ONE HUNDRED YEAR STUDY ON ARTIFICIAL INTELLIGENCE – Artificial intelligence and life in 2030. 2016 [Consultado a 22 de fevereiro de 2020]. Disponível na Internet: [https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl\\_singles.pdf](https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf)

PINA, A. S., HUSSAIN, A., ROQUE, A. C. A. – An historical overview of drug discovery. In: ROQUE, A. C. A., *Ligand-Macromolecular Interactions in Drug Discovery*, Humana Press, 2010, ISBN: 978-1-60761-243-8. p.24-29.

PRICE, W. N. – Big data and black-box medical algorithms. *Science Translational Medicine*. 10, 471 (2018).

REYMOND, J., AWALE, M. – Exploring chemical space for drug discovery using the chemical universe database. *ACS Chemical Neuroscience*. 3, 9 (2012) 649-657.

SCHANK, R. C. – What is AI, anyway?. *AI Magazine*. 8, 4 (1987) 3-13.

SEGLER, M. H. S., KOGEJ, T., TYRCHAN, C., WALLER, M. P. – Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Central Science*. 4, 1 (2018) 120-131.

SEGLER, M. H. S.; WALLER, M. P. – Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chemistry – a european journal*. 23, 25 (2017) 5966-5971.

SHAHID, M.; Li, G. – Impact of artificial intelligence in marketing: a perspective of marketing professionals of pakistan. *Global Journal of Management and Business Research: E Marketing*. 19, 2 (2019) 26-33.

SHAW, J., RUDZICZ, F., JAMIESON, T., GOLDFARB, A. – Artificial intelligence and the implementation challenge. *Journal of Medical Internet Research*. 21, 7 (2019).

SHEN, M., XIAO, Y., GOLBRAIKH, A., GOMBAR, V. K., TROPSHA, A. – Development and validation of k-nearest-neighbor QSPR models of metabolic stability of drug candidates. *Journal of Medicinal Chemistry*. 46, 14 (2003) 3013-3020.

SILVER, D., HUANG, A., MADDISON, C. J., GUEZ, A., SIFRE, L., VAN DEN DRIESSCHE, G., SCHRITTWIESER, J., ANTONOGLU, I., PANNEERSHELVAM, V., LANCTOT, M., DIELEMAN, S., GREWE, D., NHAM, J., KALCHBRENNER, N., SUTSKEVER, I., LILICRAP, T., LEACH, M., KAVUKCUOGLU, K., GRAEPEL, T., HASSABIS, D. – Mastering the game of Go with deep neural networks and tree search. *Nature*. 529, 7587 (2016) 484-489.

TURING, A. M. – Computing machinery and intelligence. *Mind*. 49 (1950) 433-460.

VAMATHEVAN, J., CLARK, D., CZODROWSKI, P., DUNHAM, I., FERRAN, E., LEE, G., LI, B., MADABHUSHI, A., SHAH, P., SPITZER, M., ZHAO, S. – Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*. 18, 6 (2019) 463-477.

YANG, M., CHEN, J., XU, L., SHI, X., ZHOU, X., XI, Z., AN, R., WANG, X. – A novel adaptive ensemble classification framework for ADME prediction. *RSC Advances*. 8, 21 (2018) 11661-11683.

ZHANG, Q., LI, J., MIDDLETON, A., BHATTACHARYA, S., CONOLLY, R. B. – Bridging the data gap from in vitro toxicity testing to chemical safety assessment through computational modeling. *Frontiers in Public Health*. 6 (2018) 1-16.

ZHAVORONKOV, A., IVANENKOV, Y. A., ALIPER, A., VESELOV, M. S., ALADINSKIY, V. A., ALADINSKAYA, A. V., TEREENTIEV, V. A., POLYKOVSKIY, D. A., KUZNETSOV, M. D., ASADULAEV, A., VOLKOV, Y., ZHOLUS, A., SHAYAKHMETOV, R. R., ZHEBRAK, A., MINAEVA, L. I., ZAGRIBELNYY, B. A., LEE, L. H., SOLL, R., MADGE, D., XING, L., GUO, T., ASPURU-GUZZIK, A. – Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*. 37, 9 (2019) 1038-1040.

ZURDO, J. – Developability assessment as an early de-risking tool for biopharmaceutical development. *Pharmaceutical Bioprocessing*. 1, 1 (2013) 29-50.