



Universidade Federal de São Paulo  
Campus Diadema



FELIPE DA SILVA LIMA

APRENDIZADO DE MÁQUINA NO CONTROLE DE  
QUALIDADE NAS INDÚSTRIAS FARMACÊUTICAS E  
ALIMENTÍCIAS

DIADEMA

2019

FELIPE DA SILVA LIMA

APRENDIZADO DE MÁQUINA NO CONTROLE DE  
QUALIDADE NAS INDÚSTRIAS FARMACÊUTICAS E  
ALIMENTÍCIAS

Trabalho de Conclusão de Curso  
apresentado como exigência parcial  
para obtenção do título de Bacharel  
em Farmácia, ao Instituto de Ciências  
Ambientais, Químicas e  
Farmacêuticas da Universidade  
Federal de São Paulo – Campus  
Diadema.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Patricia  
Santos Lopes

DIADEMA

2019

Lima, Felipe  
**Aprendizado de máquina no controle de qualidade na indústria farmacêutica e alimentícia** / Felipe da Silva Lima  
Diadema, 2019.  
41 f.

Trabalho de Conclusão de Curso (Graduação em farmácia) – Universidade Federal de São Paulo – Campus Diadema, 2019.

Orientadora: Patricia Santos Lopes

1.Aprendizado de máquina. 2.Controle de Qualidade. 3.Revisão.

CDD 658.562

FELIPE DA SILVA LIMA

APRENDIZADO DE MÁQUINA NO CONTROLE DE  
QUALIDADE NAS INDÚSTRIAS FARMACÊUTICAS E  
ALIMENTÍCIAS

Trabalho de Conclusão de Curso apresentado como exigência parcial para obtenção do título de Bacharel em Farmácia, ao Instituto de Ciências Ambientais, Químicas e Farmacêuticas da Universidade Federal de São Paulo – Campus Diadema.

Aprovado em: \_\_\_\_/\_\_\_\_/\_\_\_\_

BANCA EXAMINADORA

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Patricia Santos Lopes – UNIFESP

---

Prof<sup>o</sup>. Dr<sup>o</sup>. Newton Andréo Filho – UNIFESP

---

Me. Luís Otávio Junqueira – UNIFESP

## **AGRADECIMENTOS**

Agradeço inicialmente a minha família, por todo o apoio que me deram nos últimos anos e me deram desde o começo, me incentivando a continuar em frente com a graduação nessa universidade apesar de todas as dificuldades.

Agradeço a minha namorada Jéssica Tanaka, por ter me dado apoio e conhecimento não só para este trabalho, mas para diversos aspectos da vida. Obrigado por todos os ótimos momentos ao seu lado e por ter sido um exemplo em diversas coisas.

Agradeço também aos meus amigos que fiz nessa universidade e que espero carregalos para a vida, pois aprendi com eles, seja com o pessoal o qual dividi casa na República Bangu, seja com os amigos que estão comigo em todas as situações, sejam boas ou ruins, em jogos, festas e afins.

Agradeço a minha orientadora Prof<sup>a</sup> Patricia Lopes, por ter me proposto um desafio tão intrigante e diferente, tendo em vista as características de nosso campus e por ter sido paciente e por confiar em mim.

Agradeço em destaque a colegas e amigos que me ajudaram a ir atrás de fontes e conhecimentos da área mais distante para mim, em destaque ao Douglas Diniz Landim e Rodrigo Didier.

E à UNIFESP, por ter me proporcionado um crescimento e evolução não só na parte acadêmica, mas em todos os aspectos que tocam a minha vida. Aqui eu aprendi a não apenas ser um aluno, como aprendi a ser um ser pensante e crítico, além de um ser social que consegue contribuir para a sociedade. As experiências que tive com CAAF, Consu, Jogos Universitários, SCUDs entre outros, me agregaram em tudo.

Meus eternos agradecimentos a todos.

## RESUMO

O Controle de Qualidade é responsável pelas atividades referentes à amostragem, às especificações e aos ensaios de produtos acabados e de matérias primas, tendo vital importância nas indústrias farmacêuticas e alimentícias por estas lidarem diretamente com a saúde dos consumidores finais. Aprendizado de máquina é um subcampo de estudo da área de Inteligência Artificial que lida com agentes/algoritmos que podem melhorar o seu comportamento ou desempenho através dos estudos de suas próprias experiências prévias. Trabalhos recentes visam juntar os dois conceitos, apresentando novas técnicas e metodologias para melhorar em diversos aspectos o controle de qualidade nas indústrias farmacêuticas e alimentícias. Foram realizadas buscas de artigos nas bases de dados Scopus e Web of Science, utilizando os descritores *quality control*, *machine learning*, *pharmaceutical*, *food* e *bacterial* sendo identificados 717 estudos, dos quais foram incluídos 16 artigos completos. Desses estudos, 10 focam em análise de medicamentos e matérias primas e 6 focam em análises de alimentos. A literatura abrange diversas técnicas analíticas com diversos algoritmos de aprendizado, sendo os mais vistos a Espectroscopia Raman e FT-IR usando principalmente os algoritmos de *Support Vector Machine* e Redes Neurais Artificiais. Apesar de apresentarem bons dados estatísticos e baixo custo de aplicação, é de senso comum dos autores que ainda precisam ser feitos estudos mais aprofundados para tornar esses métodos realmente úteis e viáveis.

**Palavras-chave:** Controle de Qualidade. Aprendizado de Máquina. Indústria. Medicamentos. Alimentos. Revisão.

## ABSTRACT

Quality Control is the department responsible for the activities such as sampling, specifications and test the final products and raw material, having a vital importance at the pharmaceutical and food industries, which deal directly with the health of the final customers. Machine learning is a subfield of studies in the Artificial Intelligence field, which works with agents/algorithms that can improve their behavior or performance by studying their own previous experiences. Recent studies aim at place together both concepts, presenting us new techniques and methodologies to improve the quality control in the pharmaceutical and food industries. It was searched articles at the Scopus and Web of Science databases, using the descriptors quality control, machine learning, pharmaceutical, food and bacterial resulting in 717 articles, which 16 have been fully included in this work. Of these articles, 10 focus on drugs and raw material analysis, and 6 focus on food analysis. The literature covers several analytical techniques associated with several learning algorithms, which the most seen are Raman Spectroscopy and FT-IR using the Support Vector Machine and Artificial Neural Network algorithms. Despite of showing good statistic data and low cost application, the authors agree that still need more in-depth studies to render these methods really useful and viable.

**Keywords:** Quality Control. Machine Learning. Industry. Pharmaceutical. Food. Review.

## LISTA DE TABELAS

Tabela 1 – relação da precisão da identificação do microrganismo por algoritmo e técnica analítica estudado no trabalho de ESTELLES-LOPEZ, <i>et al.</i> , 2017.....	33
--	----

## LISTA DE ILUSTRAÇÕES

Figura 1 - Organograma padrão sugerido de uma indústria fabril. ....	13
Figura 2 - Organograma padrão sugerido da área de Garantia da Qualidade e Controle de Qualidade em uma indústria farmacêutica/alimentícia .....	14
Figura 3 - Quadro com algumas definições de Inteligência Artificial.....	15
Figura 4 - Representação de aprendizado não-supervisionado (cima) versus o de aprendizado supervisionado (baixo), usando Redes Neurais Artificiais.....	17
Figura 5 - Representação do modelo matemático de neurônios artificiais.....	18
Figura 6 - Representação de um neurônio biológico. ....	18
Figura 7 - <i>Support Vector Machine</i> (SVM).....	19
Figura 8 - Gráfico de quantidade de artigos <i>versus</i> ano usando os descritores <i>machine learning</i> e <i>pharmaceutical</i> na base Scopus.....	23
Figura 9 - Gráfico de quantidade de artigos (eixo Y) <i>versus</i> ano (eixo X) usando-se os descritores <i>machine learning</i> e <i>pharmaceutical</i> na base Web of Science. ....	23
Figura 10 - Gráfico de quantidade de artigo por país na base Scopus para os descritivos <i>machine learning</i> e <i>pharmaceutical</i> .....	24
Figura 11 - Gráfico de documento <i>versus</i> ano na base Scopus para os descritivos <i>machine learning</i> , <i>quality control</i> e <i>food</i> . ....	25
Figura 12 - Gráfico de documento <i>versus</i> ano na base Web of Science para os descritivos <i>machine learning</i> , <i>quality control</i> e <i>food</i> .....	25
Figura 13 - Gráfico de artigos por país na base Scopus pelos descritivos <i>machine learning</i> , <i>quality control</i> e <i>food</i> . ....	26
Figura 14 - maquinário de controle de qualidade <i>online</i> de ampolas desenvolvido por GE, <i>et al.</i> , 2017.....	30
Figura 15 - equipamento de leito fluído utilizado por KORTEBY, <i>et al.</i> , 2018.....	32
Figura 16 - espectrômetro utilizado no trabalho de LASCH, <i>et al.</i> , 2018. ....	33
Figura 17 - mini-espectrômetro usado no trabalho de JÄHME, <i>et al.</i> , 2016. ....	33
Figura 18 - Resumo gráfico do trabalho de ESTELLES-LOPEZ, <i>et al.</i> , 2017. ....	35

## LISTA DE ABREVIATURAS E SIGLAS

**ANVISA** - Agência Nacional de Vigilância e Sanitária  
**CELM** - Complex Extreme Learning Machine  
**ELM** - Extreme Learning Machine  
**EUA** - Estados Unidos da América  
**FT-IR** - Fourier Transformation Infrared Spectroscopy  
**GC-MS** - Gas Chromatography with Mass Spectrometer  
**HPLC** - High Performance Liquid Chromatography  
**IA** - Inteligência Artificial  
**KNN-R** - k-Nearest Neighbor Regression  
**MSI** - Multi-spectral imaging  
**OLS-R** - Ordinary Least Squares Regression  
**OS-ELM** - Online Sequential Extreme learning machine  
**PAT** - Process Analytical Technology  
**PCA** - Principal Component Analysis  
**PC-R** - Principal Component Regression  
**PLS** - Partial Least Squares  
**PLS-R** - Partial Least Squares Regression  
**QbD** - Quality by Design  
**QSAR** – Quantitative Structure-Activity Relationship  
**RF-R** - Random Forest Regression  
**RNA** – Rede Neural Artificial  
**SL-R** - Stepwise Liner Regression  
**SVM** - support vector machine  
**SVM-R** - Support Vector Machine Regression  
**UFC** - Unidades Formadoras de Colônias

## Sumário

1.	Introdução .....	12
1.1	Conceito e definição de “controle de qualidade” .....	13
1.2	Conceito e definição de “aprendizado de máquina” .....	15
1.2.1	Inteligência artificial .....	15
1.2.2	Paradigmas de aprendizado .....	16
2.	Objetivos .....	20
2.1	Objetivos gerais.....	20
2.2	Objetivos Específicos .....	20
3.	Materiais e métodos .....	21
3.1	Tipo de estudo .....	21
3.2	Amostra.....	21
3.3	Critérios de inclusão.....	21
3.4	Critérios de exclusão .....	22
4.	Resultados .....	23
5.	Discussão .....	27
6.	Conclusão .....	38
	Referências Bibliográficas .....	39

## 1. Introdução

Desde a introdução do primeiro computador doméstico, em 1977, até os dias atuais, vimos um avanço sem parar no ramo da tecnologia. Hoje em dia, temos em nossas mãos, nos *smartphones*, a mesma ou maior capacidade de processamento de dados que o sistema de computadores que levaram o homem à Lua, em 1969 (MISSION EVALUATION TEAM OF NASA MANNED SPACECRAFT CENTER, 1971)

Em maior a esse avanço da tecnologia, tivemos, dentre todos, um grande avanço na Inteligência Artificial ( RUSSEL, *et al.*, 2013). Inicialmente, uma ideia teorizada por Alan Turing ( TURING, 1950), o pai da computação, o conceito vem de que computadores poderiam evoluir à ponto de terem consciência própria, tendo noção de que existem e podendo realizar atividades que, até então, poderiam só serem consideradas para seres inteligentes evoluídos, como o ser humano. Vários filmes levaram esse conceito além, com Inteligências Artificiais super avançadas, capazes até mesmo de desenvolverem sentimentos humanos, como é visto a Skynet, do filme “Exterminador do Futuro ( The Terminator, 1984) ”, as máquinas em “Matrix ( Matrix, 1999)“ e até mesmo o David, personagem do filme “Inteligência Artificial ( A.I. Artificial Intelligence, 2001)”.

Diferente do que é extrapolado e romantizado nos filmes, a Inteligência Artificial já está presente no nosso dia-a-dia, mesmo até que não percebamos. Em forma de assistentes pessoais, nos celulares e computadores, ou até mesmo como telefonistas nos sistemas de Serviço de Atendimento ao Cliente ou telemarketing entre outros. Mas agora, o desafio é tornar os nossos sistemas produtivos mais inteligentes, seguindo essa tendência doméstica. Para isso, podemos começar de pouco em pouco a colocar sistemas baseados em Inteligência Artificial em diferentes ramos, seja diretamente na linha produtiva, na pesquisa de novos produtos ou até mesmo para a garantia e controle da qualidade dos produtos.

Este trabalho tem como foco destacar a aplicabilidade da Inteligência Artificial, focando em aprendizado de máquina, nas indústrias farmacêuticas e alimentícia que, devido as normas vigentes ( ANVISA. AGENCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2010 E AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2002) necessitam de cada vez mais melhores sistemas nos setores de garantia e controle da qualidade.

## 1.1 Conceito e definição de “controle de qualidade”

As indústrias farmacêuticas e alimentícias lidam direto com a saúde do consumidor final, tendo que ter todas as suas etapas da cadeia produtiva sendo rigidamente controladas e inspecionadas, com foco em garantir segurança e visando a proteção da saúde da população em geral, papel este direcionado à Agência Nacional de Vigilância Sanitária (ANVISA), criada e delimitada pela lei nº 9.782, de 26 de janeiro de 1999. Dentro de suas atribuições, cabem além das inspeções, tanto burocráticas quanto físicas das áreas e cadeias produtivas, regulamentar como estas podem ser feitas e delimitadas. Foi com estes objetivos que foram estabelecidas as Boas Práticas de Produção de Medicamentos ( AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2019) e de Boas Práticas de Produção de Alimentos ( AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2002), para estabelecer as diretrizes e padrões das indústrias farmacêuticas e indústrias alimentícias, respectivamente.

Em comuns nos dois documentos de Boas Práticas de Fabricação, temos o estabelecimento de “sistemas de qualidade” englobando instalações, procedimentos, processos e recursos organizacionais que garantam a qualidade do produto final, considerando desde a matérias prima e todo o processo até o produto final. Essa é a “Garantia da qualidade”, que por ser uma parte muito ampla e abrangente, é comumente subdividida em diversos setores nessas indústrias, o qual nesse trabalho será focado o “controle de qualidade”.

Figura 1 - Organograma padrão sugerido de uma indústria fabril. Fonte: VARGAS

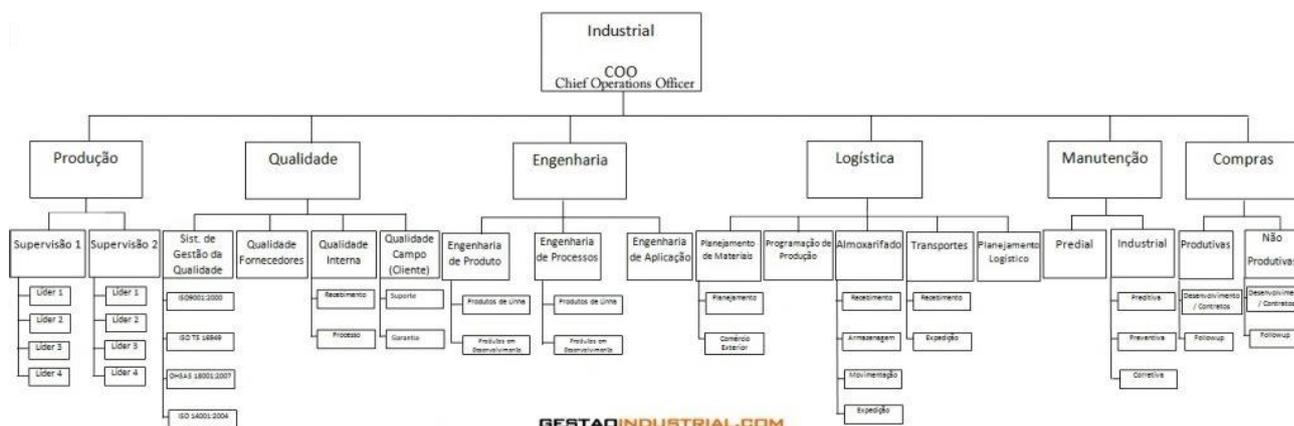
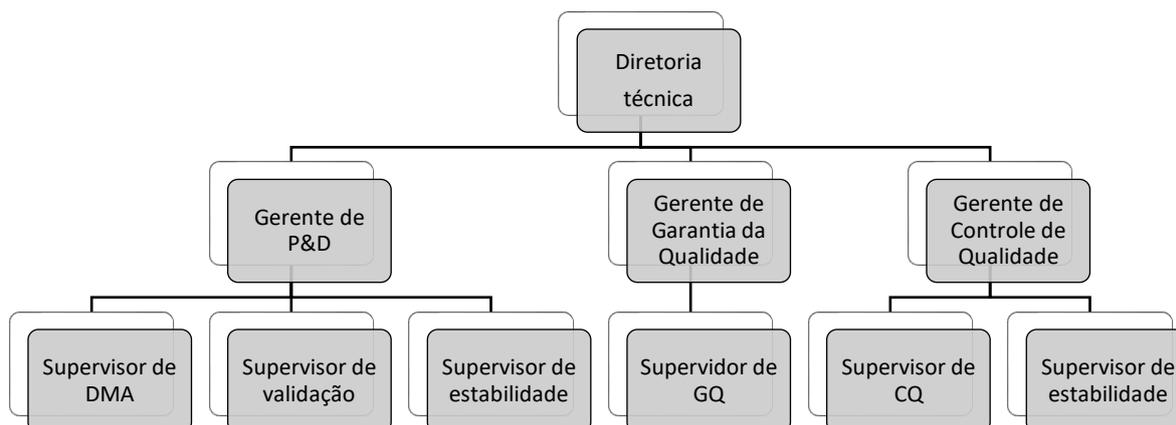


Figura 2 - Organograma padrão sugerido da área de Garantia da Qualidade e Controle de Qualidade em uma indústria farmacêutica/alimentícia



A RDC Nº 301, DE 21 DE AGOSTO DE 2019 estabelece que o Controle de Qualidade é responsável pelas atividades referentes à amostragem, às especificações e aos ensaios, bem como à organização, à documentação e aos procedimentos de liberação que garantam que os ensaios sejam executados e que os materiais e os produtos terminados não sejam aprovados até que a sua qualidade tenha sido julgada satisfatória e também que as áreas em que são empregados ensaios biológicos, microbiológicos ou de radioisótopos devem ser separadas umas das outras, assim, geralmente, na indústria farmacêutica temos a separação do Controle de Qualidade em “Físico-químico” e “Microbiológico”. Os mesmos são ditos na PORTARIA MS Nº272 1 428 DE 26 DE NOVEMBRO DE 1993, em que estabelece que “Controle de Qualidade - compreende as informações sobre os métodos e procedimentos utilizados no controle de todo o processo” através de métodos de análises para a determinação dos parâmetros que permitam verificar as características físico-químicas, os parâmetros macroscópicos, microscópicos e microbiológicos do produto, baseados em métodos internacionalmente aceitos.

Nessa mesma RDC nº 17, no Título VII, art. 570, considera que ao introduzir sistemas computadorizados, ainda deve-se seguir todas as normas de Boas Práticas de Fabricação e no inciso 1º, que não pode haver impacto na qualidade do produto.

## 1.2 Conceito e definição de “aprendizado de máquina”

### 1.2.1 Inteligência artificial

“Aprendizado de máquina” é um subcampo de estudo dentro da área de Inteligência Artificial. Para posicionar o “aprendizado de máquina” é necessário ter em mente as noções e definições de Inteligência Artificial. A tabela 1 segue com algumas definições dadas por nomes históricos no estudo da área.

Figura 3 - Quadro com algumas definições de Inteligência Artificial. Fonte: *RUSSEL, et al., 2013*

<b>Pensando como um humano</b>	<b>Pensando racionalmente</b>
<p>“O novo e interessante esforço para fazer os computadores pensarem (...) <i>máquinas com mentes</i>, no sentido total e literal.” (Haugeland, 1985)</p> <p>“[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado...” (Bellman, 1978)</p>	<p>“O estudo das faculdades mentais pelo uso de modelos computacionais.” (Charniak e McDermott, 1985)</p> <p>“O estudo das computações que tornam possível perceber, raciocinar e agir.” (Winston, 1992)</p>
<b>Agindo como seres humanos</b>	<b>Agindo racionalmente</b>
<p>“A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas.” (Kurzweil, 1990)</p> <p>“O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas pelas pessoas.” (Rich and Knight, 1991)</p>	<p>“Inteligência Computacional é o estudo do projeto de agentes inteligentes.” (Poole <i>et al.</i>, 1998)</p> <p>“AI.. está relacionada a um desempenho inteligente de artefatos.” (Nilsson, 1998)</p>

Observando o quadro da Figura 3, conseguimos analisar duas linhas de conceitos para Inteligência Artificial, seguindo outras duas vias: a de pensamento e a de ação, sendo imitando seres humanos ou por si só.

Um estudo famoso que nos ajuda também a entender como o campo da ciência da informação divide a Inteligência Artificial é o Teste de Turing. Em suma, o teste proposto por Alan Turing ( TURING, 1950) define que o computador passará no teste se um interrogador humano, depois de propor algumas perguntas por escrito, não descobrir se as respostas vêm de uma pessoa ou um computador. Para isso, o computador necessitaria de ter 4 capacidades básicas:

1. Processamento de linguagem natural para permitir que ele se comunique com sucesso em um idioma natural;
2. Representação de conhecimento para armazenar informações;
3. Raciocínio automatizado para usar as informações armazenadas com a finalidade de responder perguntas;

4. Aprendizado de máquina para se adaptar a novas circunstâncias e para detectar e extrapolar padrões.

Cada vez mais esses subcampos da Inteligência Artificial estão integrados, porém, neste trabalho, serão destacados os aspectos, assim como características, tanto matemáticas quanto lógicas, do “aprendizado de máquina” aplicados em locais específicos.

Portanto, segundo Stuart Russel e Peter Norvig (2013, Cap. 1.5) “(...)inteligência está relacionado principalmente a uma ação racional. No caso ideal, um agente inteligente adota a melhor ação possível em uma situação”.

### 1.2.2 Paradigmas de aprendizado

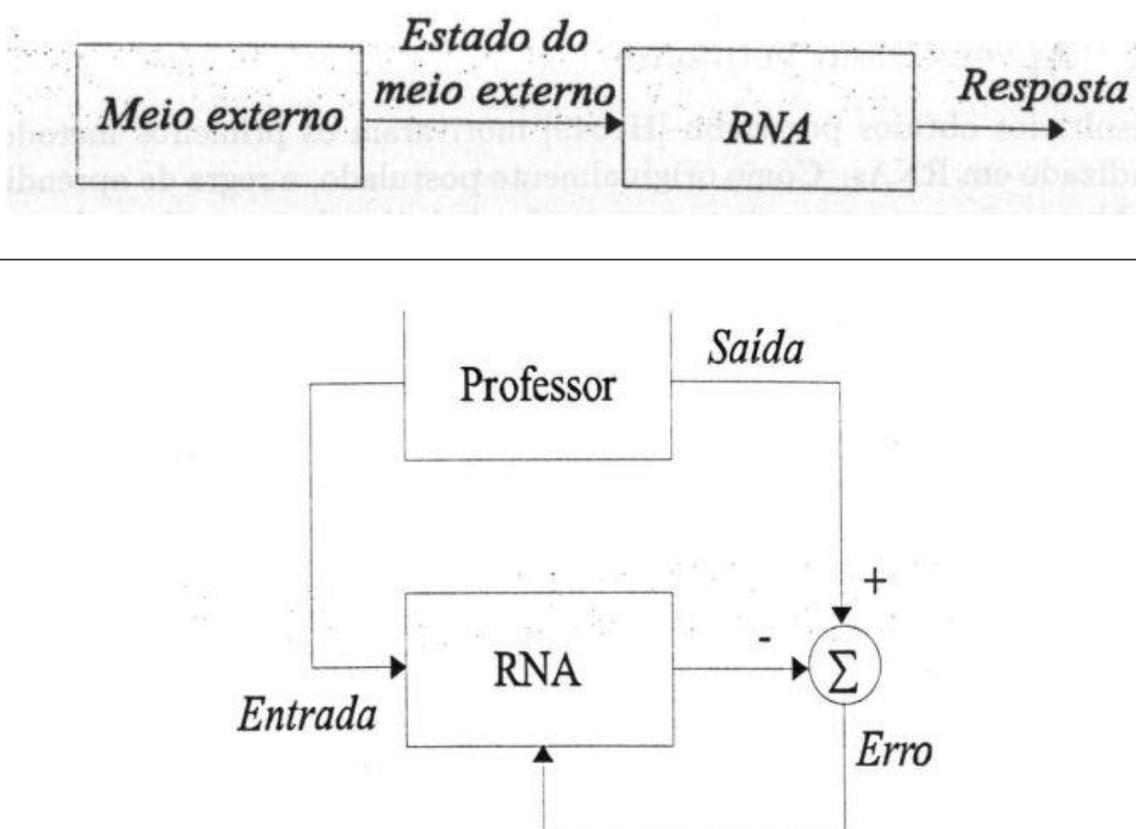
Russel e Norvig (2013, pg. 806) também estabelecem que “Um agente estará aprendendo se melhorar o seu desempenho nas tarefas futuras de aprendizagem após fazer observações sobre o mundo”. Ou seja, são agentes que podem melhorar o seu comportamento ou desempenho através dos estudos de suas próprias experiências prévias.

Para então treinar esses agentes (as máquinas), diversos métodos foram e são continuamente desenvolvidos, mas são divididos principalmente em dois paradigmas:

1. Algoritmos de **aprendizado supervisionado** são treinados (por um humano) por meio de exemplos rotulados, como uma entrada na qual a saída desejada é conhecida. Por exemplo, um equipamento poderia ter pontos de dados rotulados como “F” (falha) ou “E” (executa). O algoritmo de aprendizado recebe um conjunto de entradas junto com as saídas corretas correspondentes, e aprende ao comparar a saída real com as saídas corretas para encontrar erros. Ele, então, modifica o modelo de acordo. Através de métodos como classificação, regressão e *gradient boosting*, o aprendizado supervisionado utiliza padrões para prever os valores de rótulos em dados não-rotulados adicionais. O aprendizado supervisionado é comumente empregado em aplicações nas quais dados históricos preveem eventos futuros prováveis;
2. O **aprendizado não-supervisionado** é utilizado contra dados que não possuem rótulos históricos. A "resposta certa" não é informada ao sistema. O algoritmo deve descobrir o que está sendo mostrado. O objetivo é explorar os dados e encontrar alguma estrutura dentro deles. O aprendizado não-

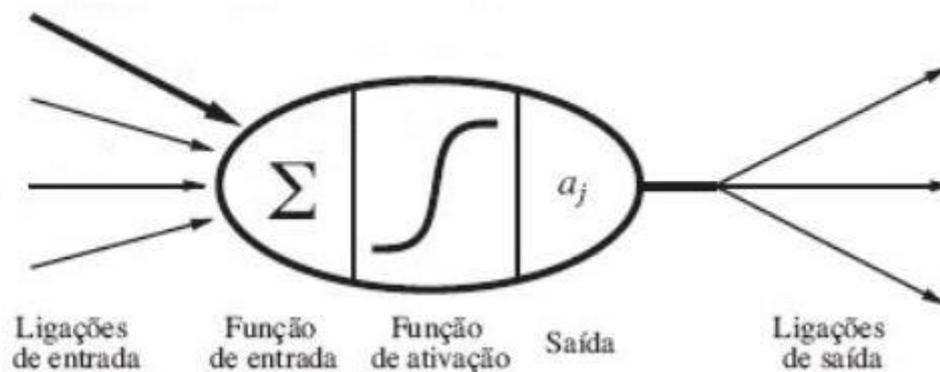
supervisionado funciona bem com dados transacionais. Por exemplo, ele pode identificar segmentos de clientes com atributos similares que podem, então, ser tratados de modo igualmente similar em campanhas de marketing; ou ele pode encontrar os principais atributos que separam segmentos distintos de clientes. Técnicas populares incluem mapas auto organizáveis, mapeamento por proximidade, agrupamento *k-means* e decomposição em valores singulares. Esses algoritmos também são utilizados para segmentar tópicos de texto, recomendar itens e identificar pontos discrepantes nos dados.

Figura 4 - Representação de aprendizado não-supervisionado (cima) *versus* o de aprendizado supervisionado (baixo), usando Redes Neurais Artificiais como exemplo. Fonte: BRAGA, et al., 2007



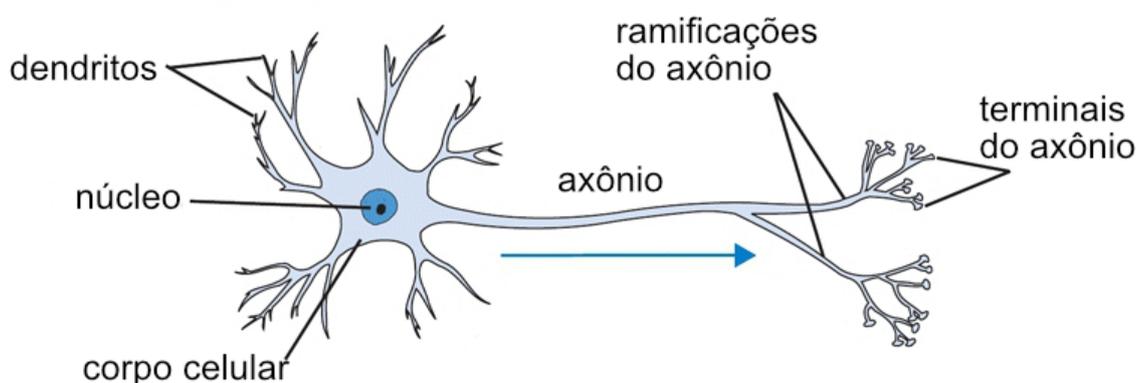
Uma técnica de aprendizado mais recente, porém que transita entre os paradigmas supracitados, é o de Redes Neurais Artificiais (RNA) ou neurônios artificiais.

Figura 5 - Representação do modelo matemático de neurônios artificiais. Fonte: *RUSSEL, et al., 2013*



Diferentes dos neurônios como células presentes no sistema nervoso, como visto na Figura 6 em comparação com a Figura 5, esses “neurônios” não são físicos e sim modelos matemáticos, porém a semelhança é que, assim como nos neurônios, há um lado de “entrada” de informações (nas células, os dendritos) e um lado de “saída” (nas células, os terminais do axônio) e assim como nas redes neurais fisiológicas, a informação percorre uma “sequência” de células, que aplicam diferentes funções e caminhos para esta, até gerar uma ação final.

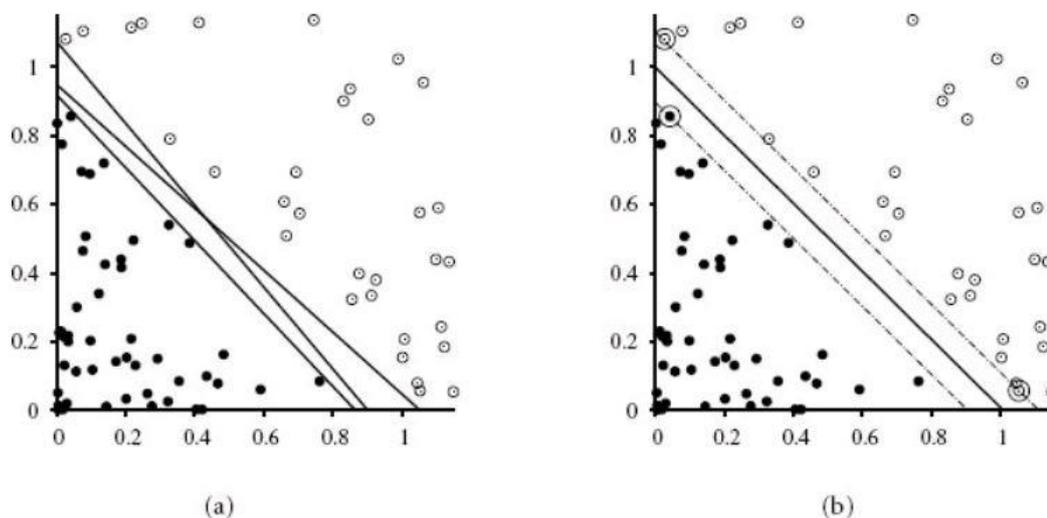
Figura 6 - Representação de um neurônio biológico. Fonte: *BEZERRA, 2016*



Um outro algoritmo bastante usado, até para os que iniciam os trabalhos na área de aprendizado de máquina, são os *Support Vector Machines* (SVM). Esse algoritmo é interessante pois constrói um separador linear em hiperplano de margem máxima, ou seja, um limite de decisão com a maior distância possível entre dois

grupos de pontos de exemplo, ajudando a generalizar com eficiência os próximos pontos de entrada. Também combinam as vantagens de modelos não paramétricos e paramétricos, tendo a flexibilidade para representar funções complexas, mas são resistentes à superadaptação. A Figura 7 exemplifica o trabalho da SVM.

Figura 7 - *Support Vector Machine (SVM)* - Em (a), duas classes de pontos e três separadores lineares candidatos. Em (b), o separador de margem máxima (linha pesada) está no ponto médio da margem. O suporte vetorial (pontos com grandes círculos) é o exemplo mais próximo do separador. Fonte: *RUSSEL, et al., 2013*



## **2. Objetivos**

### **2.1 Objetivos gerais**

Realizar uma revisão narrativa da literatura sobre a aplicabilidade de aprendizado de máquina no controle de qualidade nas indústrias farmacêuticas e alimentícias.

### **2.2 Objetivos Específicos**

- Analisar a produção científica acerca das aplicações de aprendizado de máquina em controle da qualidade nas indústrias farmacêuticas e alimentícias em escala mundial;
- Responder as perguntas de pesquisa:
  - Quais as aplicações de aprendizado de máquina em controle da qualidade em indústrias farmacêuticas e alimentícias?
  - Quais técnicas analíticas usadas em controle de qualidade nas indústrias farmacêuticas e alimentícias podem ser desenvolvidas usando aprendizado de máquina?
  - Quais as principais vantagens do uso de aprendizado de máquina em controle de qualidade?
  - Há uma alteração no interesse nesse tema nos últimos anos?
  - Quais as principais metodologias de aprendizado de máquina que podem ser aplicadas em controle de qualidade?

### 3. Materiais e métodos

#### 3.1 Tipo de estudo

Realizou-se uma revisão narrativa da literatura, que apresenta como foco uma análise qualitativa para discutir o estado da arte, ou estado do conhecimento, de um tema no período abrangido pelo trabalho. Constituída por uma análise ampla da literatura, não é necessário estabelecer uma metodologia rigorosa, como uma meta-análise, ou respostas de cunho quantitativo, no entanto, é fundamental para a aquisição e atualização do conhecimento sobre uma temática específica, evidenciando novas ideias, métodos e subtemas que têm recebido maior ou menor ênfase na literatura selecionada ( ELIAS, *et al.*, 2012). Foi então realizada uma pesquisa automática nas bibliotecas digitais **Scopus** e **Web Of Science** a partir dos descritores: *quality control*; *machine learning*; *pharmaceutical*; *food*; *bacterial*. Somente foram utilizados termos em inglês devido a ser a língua-padrão de ambos os bancos de dados.

#### 3.2 Amostra

As pesquisas foram feitas em diferentes datas abrangendo o período de abril a julho de 2019.

Foram apenas selecionados os artigos que adequavam aos critérios de inclusão, tendo então seus resumos lidos e aplicados os critérios de exclusão, resultando em 16 artigos qualificáveis, que foram então lidos na íntegra.

#### 3.3 Critérios de inclusão

Os critérios de inclusão para os artigos foram:

- Artigos publicados entre janeiro de 2009 a julho de 2019;
- Técnicas analíticas que são aplicáveis ao controle de qualidade (com reprodutibilidade e sendo quantitativa ou qualitativa com foco comparativo) tanto de medicamentos quanto de alimentos.

### 3.4 Critérios de exclusão

Os critérios para exclusão dos artigos que entram nessa revisão foram:

- Estudo de revisão de literatura;
- Artigos repetidos;
- Abrange apenas outras áreas da indústria (ex.: pesquisa e desenvolvimento; desenvolvimento analítico; etc);
- Abrange técnicas ou metodologias não aplicáveis ao controle de qualidade em indústrias farmacêuticas e alimentícias;
- Trabalha com técnica analítica extremamente específica, que não pode ser usado em outros medicamentos ou alimentos;
- Literatura de cunho reflexivo/teórico (não apresenta estudos aplicados);

#### 4. Resultados

Inicialmente, pesquisando apenas pelos descritivos *machine learning* e *pharmaceutical*, foram encontrados 717 artigos no Scopus e 291 artigos no Web of Science, com distribuição por ano mostrados nos gráficos abaixo.

Figura 8 - Gráfico de quantidade de artigos *versus* ano usando os descritores *machine learning* e *pharmaceutical* na base Scopus. Gráfico em inglês pois é gerado automaticamente pelo sistema do Scopus, padronizado em inglês.

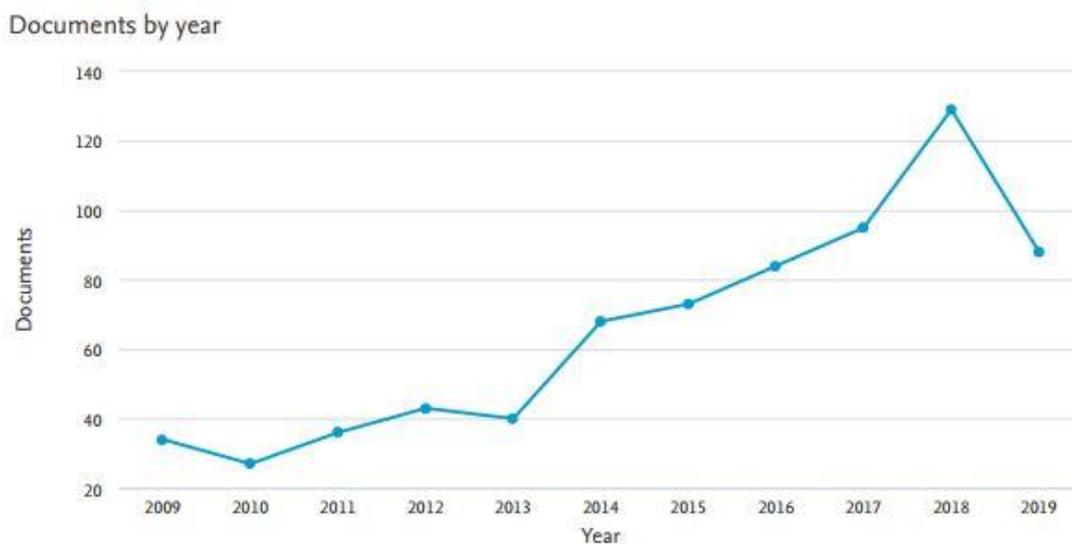
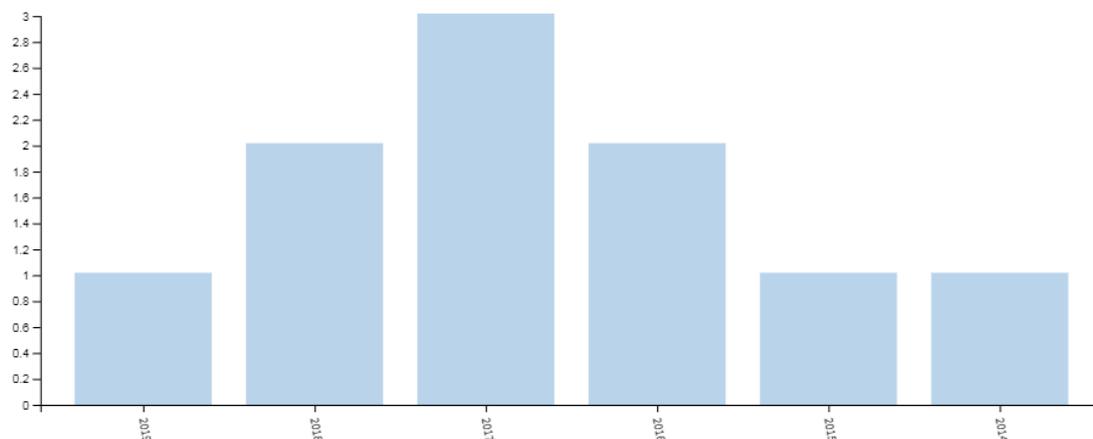
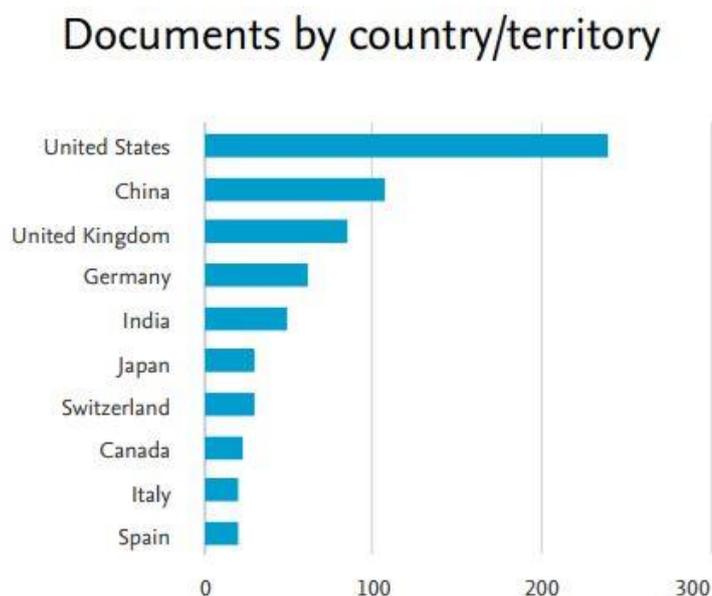


Figura 9 - Gráfico de quantidade de artigos (eixo Y) *versus* ano (eixo X) usando-se os descritores *machine learning* e *pharmaceutical* na base Web of Science (gráfico gerado automaticamente pelo sistema).



Uma vantagem da biblioteca digital Scopus é que ele fornece um gráfico de quantidades de artigos por países em que foram publicados, que nos auxilia a visualizar os países com maior influência e investimentos na área.

Figura 10 - Gráfico de quantidade de artigo por país na base Scopus para os descritivos *machine learning* e *pharmaceutical* (gerado automaticamente pelo sistema).



Foi necessário então a aplicação de mais um descritivo, então aplicou-se o termo *quality control*, resultando então em 24 artigos na base Scopus e 10 artigos na Web of Science. Estes artigos tiveram então seus títulos e resumos lidos e analisados e após aplicados os critérios de exclusão, além de eliminar artigos repetidos, resultou em 10 artigos na base Scopus e 2 na Web of Science.

Ao alterar o descritivo *pharmaceutical* para *food*, com foco em incluir a situação e necessidades da indústria alimentícia, porém, ainda mantendo os descritivos *machine learning* e *quality control*, resultou em 72 documentos na base Scopus e 25 na Web of Science, como mostram os seguintes gráficos.

Figura 11 - Gráfico de documento *versus* ano na base Scopus para os descritivos *machine learning*, *quality control* e *food* (gerado automaticamente pelo sistema).

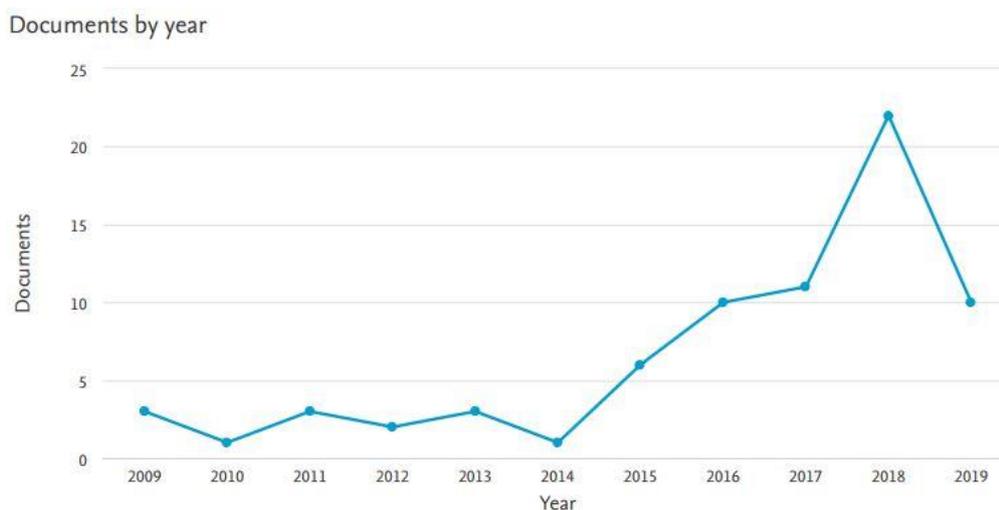
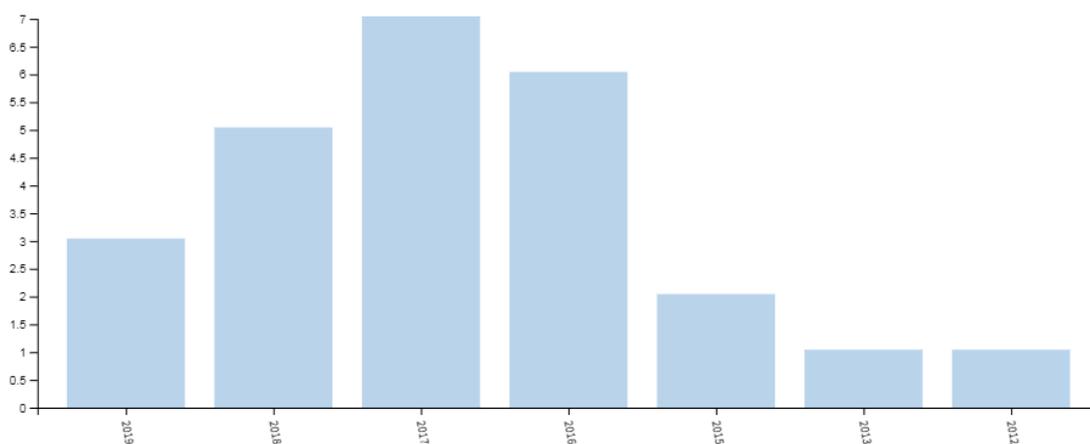
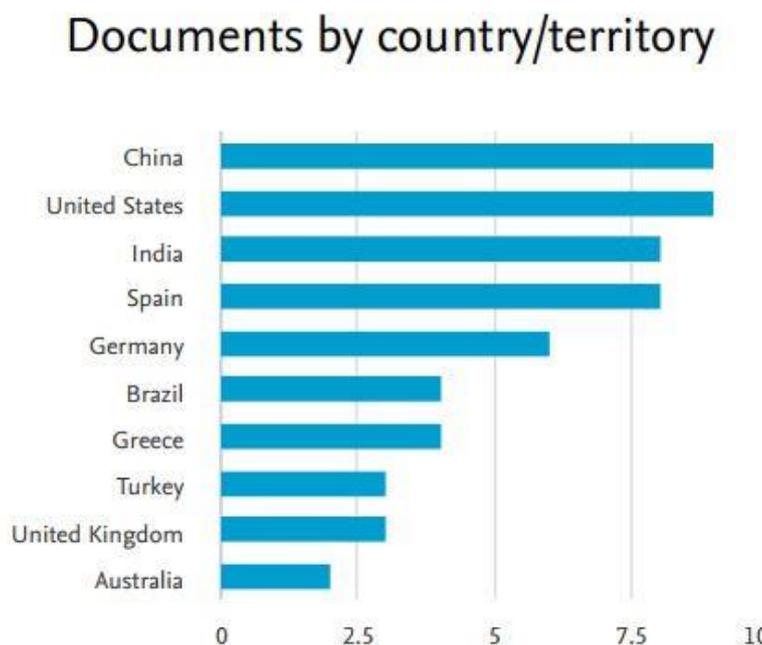


Figura 12 - Gráfico de documento *versus* ano na base Web of Science para os descritivos *machine learning*, *quality control* e *food* (gerado automaticamente pelo sistema).



Outro gráfico interessante, mais uma vez, é da base Scopus de publicação por país.

Figura 13 - Gráfico de artigos por país na base Scopus pelos descritivos *machine learning*, *quality control* e *food* (gerado automaticamente pelo sistema).



Nota-se na Figura 11 que a China publicou uma quantidade próxima de artigos que os Estados Unidos nesse período, além de aparecer o Brasil como um dos países em que mais há publicações com esses descritores.

Importante notar que, apesar dos gráficos apontarem um declínio dos artigos relacionados publicados em 2019, no momento de produção deste trabalho, o ano ainda não terminou e, portanto, estes dados podem não refletir um desinteresse no tema, como aparentam nos gráficos.

Analisando os artigos, notou-se que muitos seguiam técnicas que não eram intercambiáveis pelas indústrias de alimento para indústrias farmacêuticas, pois consistiam de técnicas para alimentos específicos ou para fatores irrelevantes para o controle de qualidade de fármacos. Foi então adicionado o descritivo *bacterial*, com o foco das análises de controle de qualidade microbiológicas, que são técnicas comuns às duas indústrias, reduzindo então o número de artigos para 6 na base Scopus e 1 na Web of Science. Após a aplicação dos critérios de exclusão, foram resultados então em 6 artigos abrangendo a área de alimento em ambos bancos de dados.

## 5. Discussão

Inicialmente, avaliando as produções científicas por país, na parte de medicamentos, podemos notar ao cruzar dados com o artigo das 10 maiores empresas farmacêuticas do mundo, escrito e publicado pela Brazil SFE Company® ( BRAZIL SFE COMPANY, 2019), que está dentro do esperado os Estados Unidos da América (EUA) serem líderes, já que 6 das 10 maiores empresas farmacêuticas do mundo são de origem e tem sede no país, além de ser o país onde mais se investe em pesquisa de medicamentos inovadores por ano. Em seguida, como reforça o artigo( BRAZIL SFE COMPANY, 2019), a China está começando a crescer nesse setor, tendo bastante medicamentos em fase pré-clínica e inicial, o que torna seu crescimento em publicações científicas relacionadas proporcionalmente crescente também. As publicações sobre controle de qualidade nesses setores crescem então proporcional ao crescimento dessas indústrias nos respectivos países, devido à necessidade de métodos que sejam cada vez mais eficientes, mais rápidos, mais amplos e de menor custo.

Quanto a produção científica na área de alimentos, também ainda sob a visão de quantidade de artigos por países, ainda podemos notar os EUA como líderes, já que, como nota-se nos dados fornecidos pela Organização Mundial do Comércio, e Divisão de Estatística da Organização para Alimentação e Agricultura das Nações Unidas ( FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS), ele é um dos maiores exportadores de alimentos do mundo. A China segue então na mesma linha por ser um dos maiores produtores de alimentos do mundo, esclarecendo a necessidade de também controlar a qualidade da entrada de alimentos importados. Também conseguimos notar nessa área uma grande influência de publicações brasileiras, pois é o segundo maior exportador de alimento do mundo e após alguns escândalos do setor que tiveram impacto mundial, em destaque, a Operação Carne Fraca da Polícia Federal( MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO, 2017), precisou-se de métodos mais modernos para se provar a qualidade do produto brasileiro, recuperando-se sua confiança, justificando o aumento no investimento de produção científica na área.

Após a leitura criteriosa dos artigos, podemos notar uma grande diversidade de técnicas analíticas somadas com diferentes algoritmos de aprendizado de máquina, porém é possível notar uma prevalência do *Support Vector Machine* (SVM), tendo este

aparecido em 8 dos 16 artigos analisados, tanto com foco na aplicabilidade deste algoritmo à técnica, quando para fins comparativos com outros algoritmos. Os autores explicam os diversos motivos para a escolha desse algoritmo, como WANG, *et al.*, 2011 justifica que os motivos da escolha foram devido ao SVM já ter grande estudo acadêmico e contestado para tarefas de classificação e regressão; habilidade para generalizar e trabalha bem com otimização dos dados até quando a quantidade das amostras é limitada. Como reforçado por WICKER, *et al.*, 2014 o SVM procura um hiperplano linear que separa as duas classes de margem maximal, porém, com dados que não podem ser separados linearmente, outras fórmulas podem ser adicionadas, como o kernel, ou outros algoritmos podem ser usados no lugar. Com isso, nota-se que, mesmo sendo um algoritmo bastante conhecido e utilizado, não há uma aplicação universal de tarefas para ele.

O segundo algoritmo que mais aparece são as Redes Neurais Artificiais (RNA), que aparecem em 3 dos 16 artigos analisados. Um resumo de como este algoritmo trabalha e sua definição pode ser lida na introdução deste trabalho (1.2.2 Paradigmas de aprendizado).

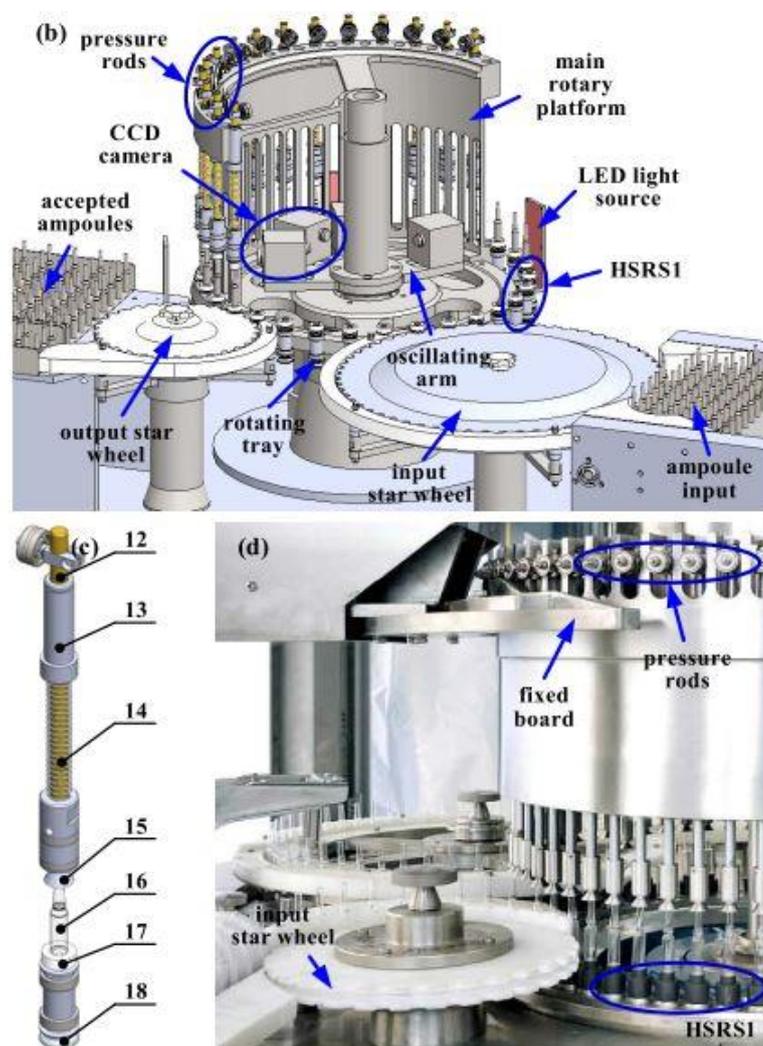
É importante ressaltar, como visto em vários destes artigos e até destacado por YIN, *et al.*, 2014 que não é necessário um computador sofisticado para se trabalhar com aprendizado de máquina, podendo usar máquinas feitas para uso doméstico e versões atualizadas do MATLAB ou até mesmo em Python, que é uma linguagem de programação de alto nível e de desenvolvimento comunitário e aberta que pode ser acessada por qualquer um sem custo obrigatório (PYTHON SOFTWARE FOUNDATION, 2019). Nota-se, portanto, que é uma tecnologia barata de ser adquirida e utilizada, podendo ser feita sem grandes custos tanto para fim de pesquisa quanto para aplicação prática, porém é extremamente necessário um profissional qualificado para trabalhar com esta ferramenta, já que, por estar em etapas iniciais de desenvolvimento e aplicação, a maioria dos algoritmos necessitam ser treinados, já que atuam por aprendizado supervisionado, como visto em 15 dos 16 artigos. Portanto, o profissional precisa não apenas entender as técnicas analíticas às quais os algoritmos estão sendo aplicadas, para saber como treinar, como também saber interagir e interpretar os dados obtidos com os algoritmos, tendo, necessariamente, algumas noções médias ou avançadas de programação e trabalho com dados complexos, tendo em vista que estes algoritmos em sua maioria ainda não estão em formas comerciais e, portanto, com uma interface interativa

desenvolvida que preza uma fácil interação. Portanto, do ponto de vista industrial, é necessário ou um profissional qualificado em áreas distintas ou uma equipe multidisciplinar, com estudos em áreas além das que são vistas atualmente nos profissionais que atuam no controle de qualidade de indústrias farmacêuticas e alimentícias, que são, em sua maioria, farmacêuticos e químicos. Porém, alguns trabalhos já estão sendo feitos com o foco de juntar a inteligência humana com a inteligência artificial, como no trabalho proposto por GAMS, *et al.*, 2014, em que ele propõe aplicar os conceitos de *Process Analytical Technology (PAT)* e *Quality by Design (QbD)* em um algoritmo de Árvore de Decisão, que resulta em destacar os impactos que certos fatores tem na linha produtiva de comprimidos, através de uma interface interativa e facilmente interpretável, necessitando que o operador tenha apenas conhecimento do processo produtivo, ou seja, a IA seria apenas um guia para que o operador consiga maximizar o processo e seguir o mais próximo possível de todos os parâmetros pré-analisados, simulando o que algumas mudanças no processo causariam de impacto no processo como um todo. Este trabalho teve os seus resultados da parte feita pela IA analisados por especialistas na área e foi verificado que o algoritmo, chamado de “*tablet classifier*”, apresentou uma precisão de 91% quando testado, ou seja, já apresenta um satisfatório grau de confiança nos dados produzidos, sendo necessário ainda, como ressaltam os autores, adicionar cada vez mais parâmetros, para se aproximar cada vez mais da realidade, mas que o programa, ao expressar os resultados em gráficos bidimensionais, ao invés de equações matemáticas complexas e desenvolvidas, como é visto no MATLAB, diminui bruscamente a barreira entre a IA e a inteligência humana.

Somando ao ponto de ser uma tecnologia de baixo custo para ser aplicada, uma outra vantagem econômica, do ponto de vista da indústria da aplicação de aprendizado de máquina, é a economia com funcionários, seja por tempo economizado com algumas tarefas menores, quanto por atividades que podem ser completamente substituídas a mão de obra humana, como é possível com o maquinário construído e estudado por GE, *et al.*, 2017. Neste trabalho, um sistema de controle de qualidade *online*, ou seja, que realiza as tarefas de detecção e padrão da qualidade ainda durante o processo produtivo, é proposto usando um sistema complexo de sensores e câmeras, como pode ser observado na Figura 13, trabalhando com um algoritmo de “*online sequential extreme learning machine (OS-ELM)*”, com foco em detectar partículas estranhas como pedaços de vidro, borracha,

cabelo e bolhas dentro de ampolas e automaticamente separando das ampolas em linha para embalagem.

Figura 14 - maquinário de controle de qualidade online de ampolas desenvolvido por GE, et al., 2017.  
Fonte: GE, et al., 2017



De acordo com o artigo, o sistema demora em torno de 0,8s por ampola para todo o processo, do momento que ele entra no sistema até a saída para continuação na linha ou no compartimento de rejeitados, para então inspeção humana mais detalhada ( GE, et al., 2017). No cenário brasileiro, esse processo de inspeção ainda pode ser encontrado feito manualmente por operários e dependendo do tamanho do lote, há necessidade de 2 ou 3 operários apenas para essa tarefa. Porém, como apontam estudos ( GUIMARÃES, et al., 2000), essa é uma atividade extremamente desgastante, tanto por constar de apenas movimentos repetitivos e desgaste ocular, por demandar grande atenção visual por parte dos operadores, que devem fazer

paradas para descanso após algumas horas de trabalhos. Isso então gera não apenas custos para a indústria, como também uma variável no controle de qualidade do processo, em que a precisão depende da experiência e treinamento do operador além do seu estado físico, que pode ser facilmente afetado por cansaço, doença ou outros fatores externos. Porém, esses fatores não impactam no maquinário proposto por GE, *et al.*, 2017, pois ele independe de operadores, tendo precisão inicial média para todas as partículas de 87,2% e erro máximo de 9,4%, valores que podem ainda ser melhorados de acordo com mais dados que podem ser obtidos pelo programa com o maior número de amostras que vão sendo analisadas, ou seja, o sistema pode ir aprendendo cada vez mais e principalmente, com um foco personalizado para um produto específico, adicionando os parâmetros específicos do produto, assim criando um método personalizado.

Também com objetivo do controle de qualidade online, o trabalho de KORTEBY, *et al.*, 2018 propõe o uso de Rede Neural Artificial para controlar a qualidade do produto obtido na granulação à quente (HMG) em um leito fluído, no equipamento visto na Figura 14, pois os fatores são mais fáceis de serem controlados e assim, pode-se calcular melhor as características do produto que serão obtidos no fim do processo. Ela determina então 3 parâmetros variáveis e após diversos produtos, usando alfa-lactose como representante de um fármaco e outros dois fundentes ligantes necessários para a técnica, ela analisa estes através de microscopia eletrônica (morfologia), difração a laser (tamanho de partícula) e a porosidade por um multipicnometro, a dureza usando um aparelho desenvolvido por eles (que informa se a deformação é viscoelástica ou elástica) e o formato das partículas usando um microscópio de varredura eletrônica, obtendo um Fator R de 0,99 entre a simulação e o resultado experimental, algo destacado pelo autor como muito bom e que é uma técnica aplicável que, acima de tudo, segue os parâmetros de QbD.

Figura 15 - equipamento de leito fluído utilizado por KORTEBY, et al., 2018. Fonte: PHOENIX EQUIPMENT CORPORATION



Voltado para o cenário de laboratórios de controle de qualidade, lendo todos os artigos, foi possível notar uma grande interação do aprendizado de máquina com as técnicas de espectroscopia, principalmente relacionado com Espectroscopia Raman, como nos trabalhos de JÄHME, *et al.*, 2016, TAHIR, *et al.*, 2017 e MARTINEZ, *et al.*, 2018 e de Espectrofotometria no Infravermelho com Transformada de Fourier (FT-IR), visto nos trabalhos de ESTELLES-LOPEZ, *et al.*, 2017; PANAGOU, *et al.*, 2010 e LASCH, *et al.*, 2018. A vantagem da escolha dessas técnicas, como em comum citado por todos os autores, é a necessidade de técnicas analíticas que sejam não-invasivas, que não destruam as amostras e rápidas, seja para foco na integridade físico-química de um medicamento que estava armazenado (TAHIR, *et al.*, 2017) como para análise da qualidade microbiológica de alimentos, baseado nos metabólitos produzidos pelos microrganismos (PANAGOU, *et al.*, 2010). Em ambas as situações, há não apenas a eliminação do tempo com preparo de amostras, como também da economia de tempo do analista de, após trabalhar os dados obtidos com o espectro para ser possível a análise dos mesmos, fazer os comparativos com valores pré-determinados.

Figura 16 - espectrômetro utilizado no trabalho de LASCH, et al., 2018. Fonte: AGILENT



Um fator interessante é que diferentes algoritmos podem ser aplicados no trabalho com os dados obtidos com esses equipamentos, como Redes Neurais Artificiais e *Partial Least Squares* para FT-IR e SVM, *Principal Component Analysis* e *k-Nearest Neighbors* para o Raman. Os valores obtidos de precisão ainda são relativamente baixos, já que estão em torno de 73,6% ( PANAGOU, et al., 2010) e 80% ( TAHIR, et al., 2017) quando apenas é analisado um tipo de amostra, mesmo que adicionado de diferentes fatores, por apenas um método. O trabalho de JÄHME, et al., 2016 propõe até o uso de um espectrômetro móvel, do modelo visto na Figura 16, com ideia de ser usado na linha produtiva, para que as análises possam ser feitas *online*. Os autores concluem que esses tipos de trabalhos ainda estão em suas fases iniciais e mais parâmetros e outros algoritmos ainda precisam ser testados, além de algum outro método que aumente a precisão dos algoritmos.

Figura 17 - mini-espectrômetro usado no trabalho de JÄHME, et al., 2016. Apresenta medidas de 148.6 x 109.2 x 46.4 mm. Fonte: OCEAN OPTICS



Com foco então na proposta de melhorar a precisão dos algoritmos de aprendizado no controle de qualidade, ESTELLES-LOPEZ, *et al.*, 2017 sugere o uso da soma dos dados obtidos de cinco técnicas analíticas diferentes, sendo estes, narizes eletrônicos; HPLC; FT-IR; cromatografia gasosa com espectro de massa (GC-MS) e *multi-spectral imaging* (MSI) e coloca sete algoritmos de aprendizado diferentes para trabalharem os dados, analisando a presença de microrganismos em carne, assim podendo comparar entre os diversos algoritmos os que apresentam as melhores precisões em cada técnica. Um resumo gráfico desse trabalho pode ser visto na Figura 17. Usando os algoritmos Ordinary Least Squares Regression (OLS-R); Stepwise Linear Regression (SL-R); Principal Component Regression (PC-R); Partial Least Squares Regression (PLS-R); Support Vector Machine Regression (SVM-R) Random Forest Regression (RF-R) e k-Nearest Neighbours' Regression (kNN-R) ela obteve diferentes precisões para cada tipo de microrganismo analisado em diferentes técnicas, como pode ser visto na Tabela 1, em que é relacionado os valores das melhores precisões para cada microrganismo, adicionado da técnica utilizada e o algoritmo que obteve essa melhor precisão.

Tabela 1 – relação da precisão da identificação do microrganismo por algoritmo e técnica analítica estudado no trabalho de ESTELLES-LOPEZ, *et al.*, 2017

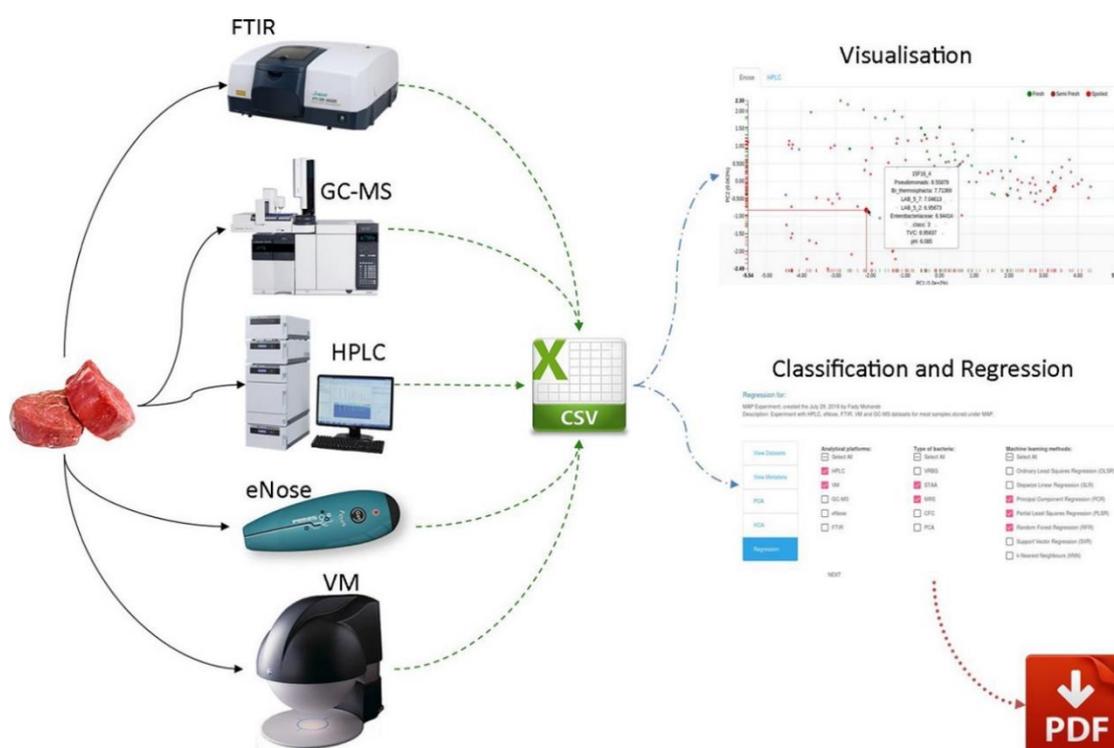
<b>Técnica</b>	<b>Microrganismos</b>	<b>Algoritmo</b>	<b>Precisão</b>
HPLC	<i>Lactobacilli</i>	RF-R	100%
HPLC	UFC total	kNN-R	93,9%
HPLC	<i>Enterobacteriaceae</i>	kNN-R	91,5%
GC-MS	<i>Enterobacteriaceae</i>	RF-R	93,9%
GC-MS	<i>Pseudomonas</i>	kNN-R	96%
Nariz eletrônico	<i>B. thermosphacta</i>	RF-R	92,8%

Os autores concluem que o foco é dar um guia para futuros trabalhos voltados para áreas específicas, ao exemplo de quais técnicas podem ser usadas para indústria e quais seriam melhores usadas por fiscais de órgãos de inspeção, tendo em vista que algumas técnicas necessitam de equipamentos fixos e outras podem ser usadas com equipamentos móveis. Ainda com foco em melhorar os dados obtidos em futuros estudos, os autores disponibilizaram os dados obtidos em uma plataforma online, via um site ([www.sorfml.com](http://www.sorfml.com)), nomeada "MeatReg". É interessante notar que as técnicas ainda poderiam ser extrapoladas da área de alimentos e serem aplicadas

na indústria farmacêutica, com foco em controle de qualidade microbiológico, em que é necessário fazer a mesma investigação de microrganismos patogênicos ou não em diferentes medicamentos.

O trabalho de ESTELLES-LOPEZ, *et al.*, 2017 ainda pode ser comparado de maneira sistemática com outros trabalhos parecidos, como o de TSAKANIKAS, *et al.*, 2016, que também segue com análise de carne, porém usando imagem multiespectral da superfície das carnes e de WANG, *et al.*, 2011, que usa nariz eletrônico com SVM para também medir a qualidade de carnes.

Figura 18 - Resumo gráfico do trabalho de ESTELLES-LOPEZ, *et al.*, 2017. Fonte: SCIENCE DIRECT



Uma outra vantagem do uso do aprendizado de máquina é tornar viável para a rotina de controle de qualidade alguns equipamentos e técnicas que não são utilizadas devido a dificuldade de se trabalhar com os dados obtidos. Isso é visto tanto no trabalho de YIN, *et al.*, 2014 e de ZHU, *et al.*, 2018. No trabalho de YIN, *et al.*, 2014, os autores testam o uso de Complex Extreme Learning Machine (CELM) em dados obtidos de um espectrômetro de terahertz pulsado. Os autores afirmam que, apesar de ser uma técnica muito útil, por permitir que façam espectros de pós sem ao menos tira-los da embalagem em que vêm, pois usa sinais de laser ultracurtos, é uma técnica ainda não muito utilizada, já que é onde se encontra o estado da arte

das pesquisas de ressonância elétron-*spin* atualmente, assim, é uma técnica ainda pouco difundida e seus dados são complexos e difíceis de se trabalhar, por serem muito maiores do que os de espectros tradicionais. Ele usa então amostras de areia, talco, sal, açúcar refinado, farinha de trigo e bicarbonato de sódio para testar os algoritmos, já que são pós que tem os espectros tradicionais muito próximos, porém, quando trabalhados com essa técnica, apresentam diferenças mais nítidas. Usando para comparar os algoritmos CELM, Extreme Learning Machine (ELM) e SVM, os autores observam que o algoritmo tem diferentes precisões ao mudar os parâmetros dos dados trabalhados, sendo que a precisão cai quanto mais se aumenta o  $\sigma$ , sendo o melhor desempenho quando se usa CELM com os parâmetros  $C=1000$  e  $\sigma=10$ , resultando em uma precisão de 99,66%. Autor concluí que o CELM é o melhor algoritmo para esse tipo de dados, pois enquanto este mantém um valor médio de precisão acima de 90%, os outros algoritmos estão muito abaixo disso. Já no trabalho de ZHU, *et al.*, 2018, a proposta é avaliar a qualidade de leite através do espectro dielétrico destes. Os autores afirmam que, atualmente, o padrão é a contagem por placa, determinando UFC/ml de leite, porém, é uma técnica que demora dias para ter o resultado, consumindo tempo e dinheiro e que o ideal seria uma técnica rápida que possa ser feita *in situ*, logo após o leite ter sido ordenhado e antes de ser enviado para processamento ou chegue no consumidor final. Usando como base trabalhos anteriores que afirmam que as atividades metabólicas dos microrganismos afetam a constante dielétrica de vários meios ( SENGUPTA, *et al.*, 2006), os pesquisadores então desenvolveram um aparelho capaz de medir a constante dielétrica do leite, de modo que possa futuramente ser adaptado para ser móvel, permitindo a análise *in situ*. A partir de análise de 150 amostras de leite e comparando com técnica clássica de contagem por placas, obteve os espectros dielétricos dos leites e aplicou em um algoritmo de SVM somado a PCA, obtendo uma precisão muito boa de 100%. Apesar de ótimos valores obtidos, autores afirmam que o trabalho ainda precisa ser mais desenvolvido antes de ser viável o uso comercial deste, já que mais parâmetros precisam ser adicionados para se chegar mais próximo da realidade em escala industrial.

Tratando a obrigatoriedade das indústrias de tratarem os seus resíduos, sejam estes químicos, biológicos ou outros ( ANVISA. AGENCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2010), garantindo o menor impacto ambiental destes, os laboratórios de controle de qualidade também monitoram os impactos ambientais na

água. Com isso em mente o trabalho de KENNICUTT, *et al.*, 2016 propõe então um modelo de cálculo, usando a técnica QSAR, se uma molécula, baseado nas suas propriedades conhecidas, vai interagir com carbono ativado ou não, permitindo então que não vá para água de dejetos caso haja essa interação, diminuindo o impacto ambiental de resíduos industriais, principalmente hormônios e ativos controlados que, mesmo em pequenos níveis de concentração, podem acarretar efeitos ou efeitos adversos na fauna, flora ou comunidades humanas que vivem ao redor de bacias ou leitos hídricos. Usou então SVM com Partial Least-Squares (PLS) e a partir de um banco de dados sobre algumas moléculas (Molecular Operating Environment), os autores concluem que obtiveram um bom equilíbrio entre performance estatística, baixo erro e pouca chance de identificar molécula errada, enquanto consegue descrever uma alta gama de elementos químicos que interagem com o carbono ativado. Os autores também desenvolveram uma ferramenta online (<http://reccr.chem.rpi.edu/WaterRF2>), com a intenção de facilitar o acesso e melhorar os dados obtidos no estudo, que ainda precisam de uma validação em bancada para ter o método mais comprovado.

Por fim, seguindo a tendência vista nos gráficos obtidos na parte de resultados e devido a serem técnicas de baixo custo financeiro de serem aplicadas e estudadas, como reforçado pelos diversos autores, mais pesquisas nessa área devem aparecer até o final do 2019, com ainda mais nos próximos anos, com este trabalho destacando todas as áreas de controle da qualidade em que pode ser aplicado o aprendizado de máquina, seja nas indústrias farmacêuticas ou de alimentos e seja no controle da matéria prima, do produto seja ainda em linha ou acabado e até mesmo os seus dejetos, facilitando e agilizando o trabalho dos diversos analistas de laboratórios envolvidos nessas etapas e partes, com diferentes técnicas que podem ter sua eficiência aumentada ou até mesmo tornando novas técnicas e equipamentos possíveis de serem aplicadas nas rotinas desses laboratórios, sempre com foco em melhorar a confiança das análises, diminuir o tempo gasto dos analistas com estas análises e diminuir os custos das análises.

## 6. Conclusão

A presente revisão narrativa da literatura apresentou os principais trabalhos dos últimos anos na área de aprendizado de máquina e suas aplicabilidades no controle de qualidade tanto de indústria farmacêuticas quanto de indústrias alimentícias, demonstrando que é um tema ainda em interesse, pretendendo destacar as principais técnicas e algoritmos que estão sendo estudados e aplicados, com foco em orientar futuros trabalhos na área.

A literatura abrange diversas técnicas, mas as mais usadas foram as análises espectroscópicas Raman e FT-IR usando principalmente algoritmos de *Support Vector Machine* e Redes Neurais Artificiais, seja para métodos de controle durante a produção ou no produto acabado. Apesar de serem métodos muito promissores, por apresentarem bons dados estatísticos e não necessitarem de grande custo, por parte das empresas, para sua aplicabilidade, a literatura segue com consenso de que ainda necessita de mais estudos na área, com dados cada vez maiores e complexos, tendo em mente a complexidade de uma rede produtiva seja farmacêutica ou alimentícia, mas que os algoritmos ainda apresentem os mesmos ou maiores graus de precisão, para se tornar algo realmente útil e viável de ser aplicado. Portanto, seguindo a tendência recente, é esperado um aumento no interesse na área, pois boa parte dos estudos são recentes e ainda há muito espaço de novas pesquisas na área ou desenvolvimento de trabalhos já realizados.

## Referências Bibliográficas

A.I. Artificial Intelligence. Direção: Steven Spielberg e Stanley Kubrick. Produção: Steven Spielberg; Jan Harlan, *et al.* Intérpretes: Haley Joel Osment. [S.l.]: Amblin Entertainment and Stanley Kubrick Productions. 2001.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. Resolução RDC nº 275, de 121 de outubro de 2002. **Portal ANVISA**, 2002. Disponível em: <[www.anvisa.gov.br/legis](http://www.anvisa.gov.br/legis)>. Acesso em: 25 Junho 2019.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. RDC Nº 301, DE 21 DE AGOSTO DE 2019. **Portal ANVISA**, 2019. Disponível em: <[http://portal.anvisa.gov.br/documents/10181/5389382/RDC\\_301\\_2019\\_.pdf/2e049461-1e8a-4bbb-8e09-8d3c04dea07d](http://portal.anvisa.gov.br/documents/10181/5389382/RDC_301_2019_.pdf/2e049461-1e8a-4bbb-8e09-8d3c04dea07d)>. Acesso em: Agosto 2019

AGILENT. Disponível em: <<https://www.agilent.com/en/products/ftir/ftir-benchtop-systems/cary-670-ftir-spectrometer>>.

ANVISA. AGENCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. Resolução da diretoria colegiada- RDC nº 17, de 16 de Abril de 2010. **Portal ANVISA**, 2010. Disponível em: <[www.anvisa.gov.br/legis](http://www.anvisa.gov.br/legis)>. Acesso em: 25 Junho 2019.

BEZERRA, S. G. T. D. A. **Reservoir Computing com Hierarquia para previsão de vazões médias diárias**. Universidade de Pernambuco. Recife. 2016.

BRAGA, A. D. P.; CARVALHO, A. P. D. L. F. D.; LUDERMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. 2ª. ed. [S.l.]: LTC, 2007.

BRAZIL SFE COMPANY. As 10 Maiores Empresas Farmacêuticas do Mundo em 2019. **Brazil SFE Company**, 2019. Disponível em: <<http://brazilsfecompany.blogspot.com/2019/04/As-10-Maiores-Empresas-Farmaceuticas-do-Mundo-em-2019.html>>. Acesso em: 9 julho 2019.

ELIAS, C. S.; AL., E. Quando chega o fim? Uma revisão narrativa sobre terminalidade do período escolar para alunos deficientes mentais. **SMAD: Revista Electrónica en Salud Mental, Alcohol y Drogas**, 2012. 48-53.

ESTELLES-LOPEZ, L. et al. An automatec ranking platform for machine learning regression models for meat spoilage prediction using multi-spectral imaging and metabolic profiling. **Food Research Internacional**, 20 Maio 2017. 206-215.  
FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. Disponível em: <<http://www.fao.org/statistics/en/>>. Acesso em: 2012.

GAMS, M. et al. Integrating Artificial and Human Intelligence into Tablet Productin Process. **PharmSciTech**, 27 junho 2014.

GE, J. et al. A system for automated detection of ampoule injection impurities. **IEEE Transaction on automation science and engineering**, 2 Abril 2017. 1119-1128.

GUIMARÃES, C. N.; MAUAD, M. L. **Estudo da atividade de "revisoras de ampolas" no controle de qualidade em indústria farmacológica**. Subdelegacia do Trabalho. Anápolis, p. 17. 2000.

JÄHME, H. et al. Recognition of pharmaceuticals with compact mini-Raman-spectrometer and automatized pattern recognition algorithms. **Optical sensing and detection**, 4 Maio 2016.

KENNICUTT, A. R. et al. A quantitative structure-activity relationship to predict efficacy of granular activated carbon adsorption to control emerging contaminants. **SAR and QSAR in Environmental Research**, 10 Agosto 2016.

KORTEBY, Y. et al. Use of machine learning tool to elucidate and characterize the growth mechanism of an in-situ fluid bed melt granulation. **Powder Technology**, 20 Março 2018. 286-295.

LASCH, P. et al. FT-IR Hyperspectral imaging and artificial neural network analysis for identification of pathogenic bacteria. **Analytical Chemistry**, 26 Junho 2018. 8896-8904.

MARTINEZ, J. C. et al. Enhanced quality control in pharmaceutical applications by combining Raman spectroscopy and machine learning techniques. **Int J Thermophys**, 7 Maio 2018. 78-91.

MATRIX. Direção: Lilly Wachowski e Lana Wachowski. Produção: Joel Silver. Intérpretes: Keanu Reeves. [S.l.]: Village Roadshow Pictures. 1999.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO, 2017. Disponível em: <<http://www.agricultura.gov.br/publicacoes/operacao-carne-fraca.pdf>>.

MISSION EVALUATION TEAM OF NASA MANNED SPACECRAFT CENTER. **Apollo 11 Mission Report**. National Aeronautics and Space Administration. Washington, D.C., p. 83 - 98. 1971. (NASA SP-238).

OCEAN OPTICS. Disponível em: <<https://oceanoptics.com/product-category/maya-series/>>.

PANAGOUE, E. Z. et al. A comparison of artificial neural networks and partial least squares modelling for the rapid detection of the microbial spoilage of beef fillets based on Fourier transform infrared spectral fingerprints. **Food Microbiology**, 16 Maio 2010. 782-790.

PHOENIX EQUIPMENT CORPORATION. Disponível em: <<https://www.phxequip.com/equipment.20716/aeromatic-strea-1-fluid-bed-dryer.aspx>>.

PYTHON SOFTWARE FOUNDATION. Python Software Foundation. **Python Software Foundation**, 2019. Disponível em: <<https://www.python.org/psf/>>. Acesso em: 31 agosto 2019.

RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. 3ª ed. ed. Rio de Janeiro: Elsevier, 2013.

SCIENCEDIRECT. Disponível em:

<<https://www.sciencedirect.com/science/article/pii/S0963996917302053>>.

SENGUPTA, S.; BATTIGELLI, D. A.; CHANG, H. C. A micro-scale multi-frequency reactance measurement technique to detect bacterial growth at low bio-particle concentrations. **Lab a Chip**, 2006. 682-692.

TAHIR, F. et al. Nondestructive evaluation of solid pharmaceutical products using wavelet transformations and multispectral data. **Journal of testing and evaluation**, 21 Novembro 2017.

THE Terminator. Direção: James Cameron. Produção: Gale Anne Hurd. Intérpretes: Arnold Schwarzenegger. [S.I.]: Hemdale and Pacific Western Productions. 1984.

TSAKANIKAS, P. et al. Exploiting multispectral imaging for non-invasive contamination assessment and mapping of meat samples. **Talanta**, 1 Setembro 2016. 606-614.

TURING, A. Computing Machinery and Intelligence. **Mind**, v. LIX, n. 236, p. 433-460, October 1950. ISSN ISSN 0026-4423.

VARGAS, R. Gestao Industrial.com. Disponível em:

<<https://gestaoindustrial.com/estrutura-organizacional/>>. Acesso em: 23 julho 2019.

WANG, D. et al. Prediction of total viable counts on chilled pork using an electronic nose combined with support vector machine. **Meat Science**, 31 Julho 2011. 373-377.

WICKER, J.; COOPER, R. Will it crystallise? Predicting crystallinity of molecular materials. **CrystEngComm**, 4 Novembro 2014. 1893-2052.

YIN, X.; HADJILOUCAS, S.; ZHANG, Y. Complex extreme learning machine applications in terahertz pulsed signals feature sets. **Computer methods and programs in biomedice**, 5 junho 2014. 387-403.

ZHU, Z. et al. A rapid method on identifying disqualified raw goat's milk based on total bacterial count by using dielectric spectra. **Journal of Food Engineering**, 20 Junho 2018. 40-51.