UNIVERSIDADE ESTADUAL DE CAMPINAS

Faculdade de Engenharia Elétrica e de Computação

GABRIEL CAUMO VAZ

APLICAÇÃO DE MODELOS DE DEEP LEARNING PARA QUALIFICAÇÃO DA ÁREA DA ENGENHARIA BIOMÉDICA : UM ESTUDO DE CASO DE VISÃO COMPUTACIONAL EM IMAGENS DE RAIO-X DA REGIÃO TORÁXICA

CAMPINAS

2023

GABRIEL CAUMO VAZ

APLICAÇÃO DE MODELOS DE DEEP LEARNING PARA QUALIFICAÇÃO DA ÁREA DA ENGENHARIA BIOMÉDICA : UM ESTUDO DE CASO DE VISÃO COMPUTACIONAL EM IMAGENS DE RAIO-X DA REGIÃO TORÁXICA

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestrado em Engenharia Elétrica, na área de TELECOMUNICAÇÕES E TELEMÁTICA.

Supervisor/Orientador: PROF. DR. YUZO IANO

Este trabalho corresponde à versão final da dissertação defendida pelo aluno Gabriel Caumo Vaz, orientada pelo(a) Prof. Dr. Yuzo Iano.

Assinatura do Orientador

CAMPINAS

2023

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

V477a

Vaz, Gabriel Caumo, 1994-
Aplicação de modelos de deep learning para qualificação da área da engenharia biomédica : um estudo de caso de visão computacional em imagens de raio-x da região toráxica / Gabriel Caumo Vaz. – Campinas, SP : [s.n.], 2023.

Orientador: Yuzo Iano.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Big data. 2. Inteligência artificial. 3. Aprendizado de máquina. 4. Aprendizado profundo. 5. Engenharia biomédica. I. Iano, Yuzo, 1950-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Informações Complementares

Título em outro idioma: Application of deep learning models for qualifying the biomedical engineering field : a case study of computational vision in x-ray images from the thoracic region
Palavras-chave em inglês:
Big data
Artificial intelligence
Machine learning
Deep learning
Biomedical engineering
Área de concentração: Telecomunicações e Telemática
Titulação: Mestre em Engenharia Elétrica
Banca examinadora:
Yuzo Iano [Orientador]
Gabriel Gomes de Oliveira
Maria Thereza de Moraes Gomes Rosa
Data de defesa: 28-02-2023
Programa de Pós-Graduação: Engenharia Elétrica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-4566-1051
- Currículo Lattes do autor: http://lattes.cnpq.br/0359200020834064

COMISSÃO JULGADORA – DISSERTAÇÃO DE MESTRADO

Candidato: Gabriel Caumo Vaz RA: 226919

Data da Defesa: 28 de fevereiro de 2023.

Título da Tese: APLICAÇÃO DE MODELOS DE DEEP LEARNING PARA QUALIFICAÇÃO DA ÁREA DA ENGENHARIA BIOMÉDICA : UM ESTUDO DE CASO DE VISÃO COMPUTACIONAL EM IMAGENS DE RAIO-X DA REGIÃO TORÁXICA

Prof. Dr. Yuzo Iano (Presidente)

Prof. Dr. Gabriel Gomes de Oliveira (Interno–Titular)

Prof. Dra. Maria Thereza de Moraes Gomes Rosa (Externo–Titular)

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de Pós-Graduação da Faculdade de Engenharia Elétrica e de Computação.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, por me proporcionar saúde, sabedoria e determinação.

Agradeço à CAPES pelo fomento financeiro ao longo do meu programa de Mestrado. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Agradeço ao meu orientador, professor Dr. Yuzo Iano, por me apoiar, confiar, estar sempre à disposição e investir seu tempo, conhecimento e experiência no meu trabalho.

À minha família, em especial meus pais, Roseli e Reinaldo, que sempre me apoiaram e acreditaram nos meus ideais.

A todos aos meus professores e amigos de curso, em especial ao ex-professor e atual amigo Rosivaldo Ferrarezi, que me apresentou ao mundo acadêmico, e aos colegas e professores Gabriel Gomes de Oliveira e Rangel Arthur, que me incentivaram a levar este trabalho até o final.

Aos meus amigos de graduação, Emerson, Guilherme e Priscila, que, apesar de seguirem caminhos diferentes, sempre me apoiaram nesse percurso.

Muito obrigado.

“Um homem que ousa desperdiçar uma hora de tempo não descobriu o valor da vida.”

Charles Darvin

RESUMO

Entre todas as áreas do conhecimento, a biológica, principalmente a medicina, sempre foi reticente no que diz respeito ao uso da tecnologia, o que é compreensível: quando se lida com seres vivos, a prudência deve ser observada e isso inclui a desconfiança com relação a resultados obtidos por computadores sem que se conheça o processo que levou a tais resultados. Contudo, com a evolução das pesquisas da área médica, em determinado momento, chegou-se num impasse: ao começar a sequenciar o genoma humano, obteve-se uma quantidade significativa de dados e, para processá-los, não havia alternativa senão os modelos computacionais.

A partir da porta aberta pela área genética, a computação se tornou cada vez mais presente na área médica, resultando na engenharia biomédica, que envolve não só as áreas cujas demandas levaram à sua criação (genética e química), mas também a medicina propriamente dita, uma vez que os profissionais da saúde têm aceitado cada vez mais o auxílio da tecnologia.

Nesse contexto, modelos de inteligência artificial, aprendizado de máquina e *deep learning* têm ganhado espaço, em especial no diagnóstico de doenças. Já existem diversas aplicações das redes neurais em outras áreas e foi questão de tempo para que as técnicas de visão computacional fossem adotadas na análise de exames de imagem (raio-x, tomografia e ultrassonografia, entre outros).

Assim, esta pesquisa tem como objetivo estudar os modelos de redes neurais com maior capacidade de analisar imagens de raio-x da região toráxica com a finalidade de identificar enfermidades específicas. Para isso, um banco de dados adequado foi selecionado, modelos de redes neurais foram estudados com base em outros casos presentes na literatura acadêmica e métricas de avaliação foram escolhidas conforme o objetivo almejado, visto que cada aplicação deve ser avaliada de forma coerente.

Para alcançar o objetivo deste trabalho, um método comparativo foi feito, por meio do treinamento convencional de uma rede neural e o treinamento de um segundo modelo com técnicas de otimização. Segundo a métrica analisada, o segundo modelo teve melhor desempenho, demonstrando que, com novas metodologias de aprendizado de máquina, melhores modelos computacionais podem ser criados, aumentando a confiança dos profissionais de saúde nessa tecnologia e melhorando a qualidade dos serviços médicos oferecidos à sociedade.

Palavras-chave: Big Data, Inteligência Artificial, Aprendizado de Máquina, Deep Learning, Engenharia Biomédica.

ABSTRACT

Among all areas of knowledge, the biological one, especially the medical science, has always been the most reticent concerning the use of technology, and it is understandable: when one copes with living organisms, prudence must be taken into account, and that includes the skepticism in results achieved by computers without knowing the process that leads to them. However, with the evolution of research in the medical field, a deadlock was eventually reached: when human genome sequencing got started, it resulted in a huge amount of data, which could not be processed without computational models.

From the moment that the genetic field opened a door, computer science has become increasingly present in the medical area, resulting in biomedical engineering, which involves not only the areas whose demands lead to its creation (genetics and chemistry), but also the medicine itself, since healthcare providers have been continuously accepting the assistance of technology.

In that context, models of artificial intelligence, machine learning, and deep learning have been gaining space, especially in the diagnosis of diseases. There are already many applications of neural networks, and it was a matter of time before the techniques of computational vision were adopted in the analysis of imaging examinations (x-ray, tomography, and ultrasonography, among others).

Therefore, this research aims to study the neural network models that have more capability of identifying specific illnesses. To this end, a suitable database was selected, neural network models were studied based on academic literature, and validation metrics were chosen according to the aimed objective, since each application must be assessed coherently.

To achieve the objective of this work, a comparative methodology was adopted by means of conventional training of a neural network and the training of a second model that used optimization methods. According to the analyzed metric, the second model had better performance, proving that, with new machine learning methodologies, better computational models can be created, increasing the confidence of healthcare providers in this technology and improving the quality of medical care provided to society.

Keywords: Big Data, Artificial Intelligence, Machine Learning, Deep Learning, Biomedical Engineering.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS, ACRÔNIMOS E SIGLAS.

CGPS - Conférence Générale des Poids et Mesures

SI - Sistema Internacional de Unidades

IoT - Internet of Things

ETL - Extract, Transform, and Load

DL - Deep Learning

DNN - Deep Neural Network

RNN - Recurrent Neural Network

CNN - Convolutional Neural Network

ANN - Artificial Neural Network

DBF - Deep Belief Network

CAP - Credit Assignment Path

PCA - Principal Component Analysis

ReLU - Rectified Linear Unit

2D - Bidimensional

3D - Tridimensional

SVM - Support Vector Machine

GMM-HMM - Gaussian Mixture Model and Hidden Markov Model

LSTM - Long Short-Term Memory

CTC - Connectionist Temporal Classification

ASR - Automatic Speech Recognition

GPU - Graphics Processing Unit

CMAC - Cerebellar Model Articulation Controller

AP - Adenocarcinoma Pulmonar

ISI - Imagem de Slides Inteiros

DCV - Doença Cardiovascular

TCBD - Tomografia Computadorizada de Baixa Dose

AUC - Area Under the Curve

RRS - Radiomic Risk Score

Sumário

1 INTRODUÇÃO

Este trabalho tem como objetivo apresentar os resultados obtidos durante uma pesquisa que abordou o uso de modelos computacionais de aprendizado de máquina profundo, também conhecido como *deep learning*, na área biomédica, mais especificamente no uso de redes neurais convolucionais em uma aplicação de visão computacional para auxiliar profissionais de saúde na detecção de doenças da região toráxica por meio de exames de raio-x.

As propostas apresentadas neste trabalho são respaldadas pelos avanços tecnológicos conquistados nos últimos anos na área de ciência de dados, seja pela disponibilidade de dados para serem analisados, pela evolução das técnicas de aprendizado de máquina, ou pelo desenvolvimento de novos *hardwares* dedicados para essas aplicações.

Para abordar o desenvolvimento desta pesquisa, este trabalho está dividido em dez capítulos.

O primeiro capítulo introduz sucintamente a dissertação aos leitores, abordando de forma genérica os conceitos envolvidos, o histórico acadêmico desta área, a metodologia aplicada, os resultados obtidos e outras informações relevantes para a compreensão deste trabalho.

O segundo capítulo tem o objetivo de contextualizar os conceitos que embasaram esta pesquisa. Por essa razão, ele está dividido em três seções: engenharia biomédica, que é, de fato, a área de atuação deste trabalho, sendo que os demais conceitos são ferramentas que auxiliaram no desenvolvimento da aplicação; *big data*, que é, basicamente, a matéria-prima desta pesquisa, visto que aplicações de aprendizado de máquina são tão eficazes quanto maior for a quantidade de dados disponíveis; *deep learning*, que é uma área da inteligência artificial extremamente avançada que tem apresentados bons resultados em diversas áreas, entre as quais, a médica.

O terceiro capítulo demonstra, por meio de dezesseis trabalhos acadêmicos, a relevância e aplicabilidade da inteligência artificial e, mais especificamente, do *deep learning* na área da saúde e biomédica, seja com aplicações clássicas, como detecção de tumores, ou mais atuais, como diagnóstico de Covid-19. A literatura acadêmica

apresenta aplicações que vão desde a análise de ruídos respiratórios até detecção de doenças em exames de imagem (raio-x e tomografia) e desenvolvimento de proteínas, demonstrando a versatilidade do aprendizado de máquina e a sua evolução nos últimos anos.

O quarto capítulo se trata da metodologia adotada durante a pesquisa e está dividido em três seções: banco de dados, na qual é apresentado o conjunto de imagens que foi selecionado para desenvolver o modelo de *deep learning* (ChestX-ray14) e os critérios adotados na sua escolha; arquitetura e métricas, na qual são descritas a construção da rede neural, bem como a função de cada uma das camadas que a compõe, e as métricas de avaliação adotadas para determinar se o modelo atingiu o objetivo ao final da otimização; treinamento, na qual são apresentados o *hardware* e o *software* utilizados durante a pesquisa e os detalhes envolvidos no processo de otimização da rede neural.

O quinto capítulo traz a análise dos resultados. Durante a pesquisa, dois modelos de rede neural foram treinados, um para obter uma referência comparativa e outro que é a aplicação almejada pela pesquisa. Por essa razão, esse capítulo está dividido em três partes, sendo a primeira e a segunda dedicadas à análise individual de cada modelo com base nas métricas estabelecidas no capítulo 4 e a terceira dedicada à comparação dos modelos, evidenciando as vantagens da rede neural proposta. Todos os dados utilizados na análise são apresentados em forma de gráficos.

No sexto capítulo, é feita uma discussão com base nos resultados apresentados no capítulo cinco a fim de evidenciar as vantagens do uso de *deep learning* e *big data* na área médica. Também se discute sobre a evolução tecnológica que permitiu o desenvolvimento desta pesquisa e se apresenta uma visão da relação complementar entre humanos e modelos computacionais, principalmente o *deep learning*.

O sétimo capítulo se refere à conclusão do trabalho, que é de suma importância, pois, a partir dela, pode-se pensar em trabalhos futuros que sigam a mesma métrica empregada nesta dissertação.

O oitavo capítulo apresenta propostas de trabalhos futuros a serem considerados no programa de doutorado pela Unicamp.

Posteriormente, são inseridas as referências bibliográficas utilizadas durante a pesquisa e, por fim, são apresentados os anexos que apoiaram esta dissertação, que contêm o algoritmo utilizado no treinamento das redes neurais e o processo de treinamento em si.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo abordará os principais conceitos que embasam esta pesquisa, quais sejam, engenharia biomédica*, Big Data e Deep Learning*, no que dizem respeito, respectivamente, à área de aplicação do estudo, à matéria-prima da pesquisa e às ferramentas utilizadas.

2.1 Engenharia biomédica

A engenharia biomédica é uma área que une os princípios das ciências exatas e da saúde com a finalidade de buscar inovações tecnológicas que possam ser aplicadas na medicina, o que envolve métodos de prevenção (monitoramento de doenças infecciosas), diagnóstico (instrumentos ou sistemas que auxiliem na detecção de doenças) e terapia (desenvolvimento de fórmulas, compostos ou equipamentos que facilitem a recuperação de um paciente) (FONSECA, CARDOSO, *et al.*, 2019) (GABRIEL, QUARESMA, *et al.*, 2018). Os engenheiros biomédicos são aptos a ocupar cargos em hospitais, clínicas, empresas e universidades, podendo atuar como administradores hospitalares, engenheiros clínicos, desenvolvedores de novos equipamentos e pesquisadores. Pela integração entre campos do conhecimento, a engenharia biomédica é considerada uma área multidisciplinar que envolve conceitos da química, física, biologia, medicina, engenharia elétrica, engenharia mecânica, ciência dos materiais e computação (ENDERLE, 2012).

Nas subseções a seguir, serão apresentadas aplicações práticas da engenharia biomédica.

2.1.1 Definição da Engenharia Biomédica.

Muitos dos problemas enfrentados pelos profissionais de saúde na atualidade são extremamente significativos para os engenheiros porque envolvem os aspectos fundamentais da análise de dispositivos e sistemas (concepção e aplicação prática), que estão no cerne de processos que são fundamentais para a prática da engenharia. Os problemas de concepção medicamente relevantes podem variar desde construções muito complexas em grande escala, tais como a concepção e implementação de laboratórios clínicos automatizados, instalações de rastreio multifásico (centros que permitem a realização de muitos testes) e sistemas de informação hospitalar, até a criação de dispositivos relativamente pequenos e simples,

tais como eletrodos e transdutores, que podem ser utilizados para monitorar a atividade de processos fisiológicos específicos, seja num ambiente de investigação ou num ambiente clínico. Essas aplicações abrangem as muitas complexidades da monitoração remota e da telemetria e incluem os requisitos de veículos de emergência, salas de operações e unidades de cuidados intensivos (BRONZINO, 2005).

O sistema de cuidados de saúde americano, por exemplo, engloba muitos problemas que representam desafios para os membros da classe de engenharia denominados engenheiros biomédicos. Uma vez que a engenharia biomédica envolve a aplicação de conceitos, deve-se ter domínio de praticamente todas as áreas de engenharia (por exemplo, elétrica, mecânica e química) para resolver problemas específicos relacionados com os cuidados de saúde. As oportunidades de interação entre engenheiros e profissionais de saúde são muitas e variadas (BRONZINO, 2005).

Os engenheiros biomédicos podem envolver-se, por exemplo, na concepção de uma nova modalidade de imagiologia médica ou no desenvolvimento de novos dispositivos médicos protéticos para ajudar pessoas com deficiências. Mesmo que a abrangência da engenharia biomédica seja considerada bastante clara, há muitas opiniões contraditórias relativas à sua definição devido à diversidade de termos, como, por exemplo, engenharia biomédica, bioengenharia, engenharia biológica e engenharia clínica ou médica, que são reconhecidos por diretórios educacionais (BRONZINO, 2005).

Embora uma parcela dos profissionais tenha definido a bioengenharia como o termo geral utilizado para descrever todo este campo, ela é geralmente definida como uma atividade básica orientada para a investigação estreitamente relacionada com a biotecnologia e a engenharia genética, ou seja, a modificação de células animais ou vegetais ou partes de células para melhorar plantas ou animais ou para desenvolver novos microrganismos para fins benéficos. Na indústria alimentar, por exemplo, isto tem significado a melhoria de estirpes de leveduras para fermentação. Na agricultura, os bioengenheiros podem estar preocupados com a melhoria do rendimento das culturas através do tratamento de plantas com organismos ou da redução dos danos causados pelas geadas (BRONZINO, 2005).

É evidente que a bioengenharia será essencial para o futuro, tendo impacto na qualidade de vida humana. O pleno potencial desta especialidade é difícil para imaginar. As atividades típicas incluem: o desenvolvimento de espécies melhoradas de plantas e animais para a produção de alimentos; invenção de novos testes de diagnóstico médico de doenças; produção de vacinas sintéticas a partir de células clonais; engenharia bioambiental para proteger a vida humana, animal e vegetal de substâncias tóxicas e poluentes; estudo das interações proteína-superfície; modelação da cinética de crescimento de células de levedura e hibridoma; investigação em tecnologia enzimática imobilizada; desenvolvimento de proteínas terapêuticas e anticorpos monoclonais (BRONZINO, 2005).

Dessa forma, evidencia-se a ampla magnitude e importância desta área para distintos aspectos que tange a sociedade, que nada mais é do que o avanço significativo da ciência para qualificar as técnicas adotadas atualmente, possibilitando, assim, um avanço significativo na saúde e qualidade de vida.

2.1.2 Funcionamento do corpo humano

As aplicações voltadas para a área da saúde dependem fortemente da compreensão do funcionamento do corpo humana e a engenharia biomédica não seria diferente. A partir deste ponto de vista, pode-se aplicar técnicas da engenharia para a obtenção de informações sobre as transformações químicas que ocorrem no corpo humano envolvendo alimentos e oxigênio, por exemplo, ou sobre as propriedades dos diferentes tecidos do corpo humano. Tais informações são essenciais para que se possa modelar o funcionamento do corpo humano e prever os efeitos de compostos terapêuticos (RONDON, DA SILVA, *et al.*, 2018).

2.1.3 Biomateriais e desenvolvimento de próteses

Há uma série de situações que podem levar uma pessoa a precisar de próteses, tais como, acidentes, doenças degenerativas e deficiências físicas. Do ponto de vista funcional, as próteses devem ser capazes de substituir de forma minimamente satisfatória partes do corpo humano, ou seja, se uma pessoa precisa de uma prótese para substituir uma perna, por exemplo, tal prótese deve permitir que a pessoa se locomova, mesmo que de uma forma limitada. Do ponto de vista biológico, as próteses devem ser desenvolvidas com materiais compatíveis com o corpo humano e existem

vários parâmetros, tais como a esterilidade do biomaterial, que devem ser observados para garantir a biocompatibilidade, a fim de minimizar a possibilidade de rejeição por parte do corpo humano (ANTUNES, DO VALE, *et al.*, 2002).

2.1.4 Bioinformática e diagnósticos

Nos últimos anos, a quantidade de dados referentes à área médica tem aumentado exponencialmente. Embora a disponibilidade de um volume maior de dados seja benéfica para a engenharia biomédica, essas informações devem ser tratadas e processadas por computadores e é nesse aspecto que a bioinformática ganha relevância. Uma parte dos sistemas desenvolvidos dentro dessa área é dedicada ao diagnóstico de enfermidades, seja por mapeamento do genoma humano, por análise de dados numéricos (um hemograma, por exemplo) ou por processamento de imagens, o que inclui a análise de radiografias, tomografias, ultrassonografias, ressonâncias.

2.2 Big Data

Por definição, *Big Data* são grandes conjuntos de dados cuja origem remonta às décadas de 1960 e 1970, quando os primeiros *data centers* foram instalados. Com o passar dos anos, principalmente a partir dos anos 2000, os dados armazenados passaram a ter maior variedade, volume e velocidade (tríade conhecida como três Vs do *Big Data*) (ORACLE, 2021).

Em outras palavras, *Big Data* é uma evolução dos conjuntos de dados conhecidos até o final do século XX. Tal evolução diz respeito aos tipos (variedade) dos dados, à quantidade (volume) de informação e à velocidade com a qual os dados podem ser transmitidos. O *Big Data*, no entanto, não implica numa evolução estrita dos sistemas de armazenamento e transmissão, uma vez que sistemas de processamento de dados convencionais não são capazes de lidar com essa nova realidade. Novas técnicas precisaram e continuam a ser desenvolvidas para essa finalidade e, na maioria das vezes, são utilizados algoritmos de aprendizado de máquina para resolver problemas que não podiam ser resolvidos antes. A Figura 1 apresenta uma nuvem de palavras com os principais conceitos relacionados com *Big Data*.

Figura 1: O que é *Big Data*

Fonte: (ACESSA.COM , 2018)

2.2.1 Importância do *Big Data*

No que diz respeito ao conceito de *Big Data*, mais importante do que ter uma grande quantidade de dados armazenados é saber o que fazer com eles. Do ponto de vista empresarial, pode-se obter dados de fontes diversas e analisá-los para encontrar soluções que permitam redução de custos, economia de tempo, desenvolvimento de novos produtos, otimização de ofertas e tomada de decisões mais inteligentes (TAURION, 2013) (GOES, 2014).

Outras aplicações possíveis quando se combina *Big Data* com técnicas adequadas para análise de dados estão presentes em ambientes industriais (determinação em tempo real das causas de erros, problemas e defeitos), comerciais (detecção de hábitos de compras para sistemas de recomendação) e financeiros (cálculo de carteira de risco e detecção de fraudes) (VEGA, ORTEGA e JOYANES-AGUILAR, 2015).

2.2.2 Os três Vs do *Big Data*

Conforme apontado anteriormente, os três Vs do *Big Data* significam variedade, volume e velocidade. Variedade diz respeito aos diferentes tipos de dados disponíveis. Quando os primeiros *data centers* foram instalados, os dados eram perfeitamente estruturados, de forma que se adequassem a bancos de dados relacionais,

construídos com associações de tabelas. Contudo, conforme os dados evoluíam para se tornarem o que hoje é conhecido como *Big Data*, eles se tornaram menos estruturados, levando até mesmo à criação de bancos de dados não relacionais, a fim de flexibilizar a forma de armazenamento dos novos dados. A complexidade desse novo modelo de manipulação de dados é tamanha que se exige um pré-processamento para extrair significado dos metadados (TOLSTOY, 2021).

O volume dos dados diz respeito à quantidade de informação armazenada. Essas informações podem ter origem nas mais diversas fontes, como redes sociais, páginas na Internet, aplicativos para dispositivos móveis, e dispositivos equipados com sensores (principalmente com o advento da Internet das Coisas), entre outros. Do ponto de vista quantitativo, as empresas possuem bancos de dados da ordem de dezenas de *terabytes* até centenas de *petabytes* (KING, 2013). A tendência de crescimento do volume de dados armazenados digitalmente é tão alta que, na 27ª reunião da Conferência Geral de Pesos e Medidas (CGPS, do francês *Conférence Générale des Poids et Mesures*), realizada entre os dias 15 e 18 de novembro de 2022, foi aprovada uma resolução que cria quatro novos prefixos no Sistema Internacional de Unidades (SI), sendo dois múltiplos e dois submúltiplos, face à "necessidade da ciência dos dados de, num futuro próximo, expressar quantidades de informação digital usando ordens de magnitude superiores a 10^{24}." Com isso, abre-se a possibilidade de utilizar os termos *ronnabytes* para representar 10^{27} *bytes* e *quettabytes* para representar 10^{30} *bytes* (BUREAU INTERNATIONAL DES POIDS ET MESURES, 2022).

A velocidade diz respeito a quão rápido os dados são recebidos e gerenciados. Esse aspecto diz respeito tanto à velocidade de transmissão dos dados, uma questão amplamente abordada quando se fala em 5G, quanto à velocidade de processamento. Em geral, quanto maior for a necessidade de avaliações em tempo real, menor é a tendência à gravação de dados no disco rígido durante o processamento, uma vez que a transferência de dados para a memória tem menor latência (BATTY, 2016).

Embora a literatura se refira, na maioria das vezes, aos três Vs do *Big Data*, nos últimos anos, tem-se falado em dois Vs adicionais, que significam valor e veracidade. O valor intrínseco de um dado só tem utilidade quando ele é descoberto e, para isso, não basta simplesmente aplicar técnicas de análise de dados, afinal, tais técnicas são capazes de extrair características dos dados, mas não fazem juízo de

valor das características detectadas. Essa interpretação cabe aos especialistas da área na qual a análise for aplicada (ORACLE, 2021).

A veracidade tem relação com a confiabilidade dos dados, ou seja, o quanto os dados analisados representam a realidade. Com o *Big Data*, dada a maior quantidade de informações, tende-se a obter respostas mais completas, o que implica numa maior confiança nos dados e representa uma abordagem inovadora no tratamento de problemas dependentes da análise de dados (OHM, 2012) (BUGHIN, 2016) (WANG, 2018) (RAGUSEO, 2018).

2.2.3 *Big Data* na prática

O *Big Data* se tornou algo intrínseco nas empresas e na sociedade em geral. Para compreender isso, basta analisar o histórico das empresas com maior valor de mercado da atualidade. Enquanto, em 2011, das cinco empresas mais valiosas do mundo, três eram do setor energético, em 2021, a lista passou a ser liderada exclusivamente por empresas do setor tecnológico (INSPER INSTITUTO DE ENSINO E PESQUISA, 2022). Grande parte do valor agregado que oferecem é proveniente de seus dados, que são analisados continuamente para que levem a uma maior eficiência para as empresas e guiem o desenvolvimento de novos produtos (ZIBIN ZHENG, 2013).

Avanços tecnológicos na área de armazenamento de dados, inclusive com estruturas de código aberto, diminuíram consideravelmente o custo da computação de dados, o que impulsionou o desenvolvimento de sistemas capazes de lidar com o *Big Data*, tornando o modelo de negócio mais lucrativo (SINANC, 2013) (L. GU, 2013).

Uma parte importante do *Big Data* é a fonte dos dados. De fato, os dados disponíveis para sistemas de computação só atingiram volumes e fluxos característicos do *Big Data* com o advento das redes sociais, quando os usuários passaram a gerar continuamente uma grande quantidade de informações. Esse efeito não só continua a ser verdadeiro, como se tornou ainda mais acentuado, com a criação e difusão das plataformas de *streaming* e *e-commerce* (um fato que remete novamente às empresas com maior valor de mercado da atualidade). Contudo, apesar dos usuários de serviços disponibilizados na Internet ainda serem uma fonte importante de dados, esse papel deixou de ser exclusivamente humano (BARNES, 2013) (BATISTIč, 2019).

A Internet das Coisas (IoT – do inglês *Internet of Things*) é caracterizada por uma constelação de dispositivos (geralmente equipados com sensores para as mais diversas finalidades) interconectados entre si e com a rede mundial de computadores. Logo, com a expansão desse mercado, há cada vez mais objetos e dispositivos conectados à Internet, de forma que criam fluxos de informações que, por sua vez, alimentam sistemas de armazenamento e análise de dados. A Internet das Coisas é um exemplo clássico dos elementos característicos do *Big Data* (variedade, volume e velocidade) (MANNING, 2013) (GUTMANN, 2018) (EIJNATTEN, 2013).

Apesar da evolução do conceito *Big Data*, ainda há margem para crescimento, visto que a computação em nuvem expandiu ainda mais os horizontes do *Big Data*. A computação em nuvem oferece ao desenvolvedor uma liberdade maior na escalabilidade dos sistemas, afinal, eles podem simplesmente criar *clusters ad hoc* para testar um subconjunto de dados ao invés de precisar adquirir equipamentos físicos dedicados para a execução dos testes (WU, 2016) (JIFA, 2014).

2.2.4 Aplicações do *Big Data*

O *Big Data* pode auxiliar no gerenciamento de várias atividades de negócios, desde a experiência do cliente até análises avançadas. Seguem abaixo alguns exemplos.

2.2.4.1 Desenvolvimento de produtos

Pode-se usar o *Big Data* para prever as demandas dos clientes por meio de modelos preditivos capazes de identificar os principais atributos dos produtos e serviços pré-existentes e modelar a relação entre tais atributos e o sucesso de vendas. Com base nessa análise, as empresas podem desenvolver novos produtos com maiores chances de agradar os consumidores. Outra possibilidade é uma análise de mercado que leve em consideração o público alvo de uma empresa com a finalidade de determinar o melhor local para a instalação de uma nova loja e quais produtos essa loja deve ter no estoque, dada a maior probabilidade de consumo na região (LI, 2015) (ZHAN, 2018) (JOHNSON, 2017).

2.2.4.2 Manutenção preditiva

Empresas que contam com linhas de produção têm uma preocupação constante com a possibilidade de falhas mecânicas nos seus equipamentos. Tais

falhas podem ser previstas e corrigidas antes que a máquina se deteriore a ponto de precisar de uma manutenção corretiva. Quando se programa uma manutenção com base em dados estruturados, ou seja, critérios técnicos que são, geralmente, fornecidos pelo fabricante do equipamento, diz-se que a manutenção é preventiva. Contudo, com a expansão de sistemas de automação, dados não estruturados, provenientes de sensores, passaram a fazer parte dessa previsão de falhas e, neste caso, diz-se que a manutenção é preditiva, que se baseia nas condições em tempo real do equipamento e tem como objetivo reduzir a necessidade de manutenções preventivas e, principalmente, corretivas. Ao obter indicações de problemas potenciais antes que eles surjam, pode-se implementar a manutenção de maneira mais econômica e maximizar a disponibilidade de peças e equipamentos (DAILY, 2017) (YAN, 2017).

2.2.4.3 Experiência do cliente

O principal ativo de uma empresa são os seus clientes e, portanto, faz-se necessária uma visão clara da experiência do cliente com os produtos adquiridos e com a empresa em si. Essa percepção nunca foi tão tangível como hoje, uma vez que o *Big Data* permite reunir dados de redes sociais, visitas a páginas na Internet, e registros de chamadas, entre outras fontes, para melhorar a interação e maximizar o valor entregue, oferecendo ofertas e atendimento personalizados (SPIESS, 2014) (HOLMLUND, 2020) (DIAZ-AVILES, 2015).

2.2.4.4 Detecção de fraudes

A detecção de fraudes está intimamente relacionada com aplicações financeiras, sejam bancos ou seguradoras. Com o aumento exponencial de dados disponíveis (*Big Data*) e evolução dos sistemas computacionais, o que permite o processamento dos dados em tempo real, as instituições financeiras são capazes de prever, com precisão cada vez maior, os casos de fraudes, possibilitando o bloqueio da operação considerada fraudulenta (CARDENAS, 2013) (TRIERWEILER, 2019).

2.2.4.5 Machine learning

No cerne de toda a evolução proporcionada pelo *Big Data*, está o aprendizado de máquina. A relação entre esses dois conceitos é praticamente simbiótica, uma vez que os dados sem um modelo de processamento adequado não têm valor real e o

aprendizado de máquina sem dados não é funcional. Ao contrário dos modelos de computação tradicionais, baseados em algoritmos pré-determinados, agora pode-se fazer com que as máquinas processem os dados sem serem explicitamente programadas. Em outras palavras, a máquina adquire a capacidade de aprender a lidar com os dados (QIU, 2016) (ZHOU, 2017) (L'HEUREUX, 2017) (BEAM, 2018).

2.2.5 Desafios do *Big Data*

Embora novas tecnologias de armazenamento de dados tenham sido desenvolvidas, estima-se que os volumes de dados tenham um aumento exponencial, com uma razão de duas vezes a cada dois anos. Portanto, há uma luta constante por parte das empresas para acompanhar as mudanças em seus dados e encontrar maneiras de armazená-los com eficiência (LABRINIDIS, 2012) (MARX, 2013) (JAGADISH, 2014).

Contudo, arquivar dados brutos não é suficiente, pois, para que sejam úteis numa análise, eles devem ser tratados a fim de proporcionar clareza e organização para uma análise válida. Esse processo é comumente chamado de pré-processamento e os cientistas de dados gastam de 50% a 80% de seu tempo preparando os dados para que possam ser utilizados (LABRINIDIS, 2012) (MARX, 2013) (JAGADISH, 2014).

Conforme mencionado anteriormente, o *Big Data* está relacionado com um grande volume de dados que, ao contrário dos bancos de dados tradicionais, podem ser não estruturados. No entanto, não se deve confundir dados não estruturados com dados não organizados. Essa organização se dá com algumas etapas, quais sejam, integração, gerenciamento e análise.

A integração está relacionada com o fato de que o *Big Data* coleta dados a partir de diferentes fontes e as ferramentas tradicionais de extração, transformação e carregamento (*Extract, Transform, and Load* - ETL) geralmente não desempenham bem essa tarefa. Há necessidade de novas estratégias e tecnologias para analisar grandes conjuntos de dados em escala de *terabyte* ou mesmo *petabyte*. Durante a integração, é necessário coletar os dados, processá-los e garantir sua formatação e disponibilidade para que os analistas possam utilizá-los (LOHR, 2012) (CUI, 2016) (O'MALLEY, 2014).

O gerenciamento diz respeito à estratégia de armazenamento dos dados, visto que o grande volume de dados é um pressuposto do *Big Data*. As opções para

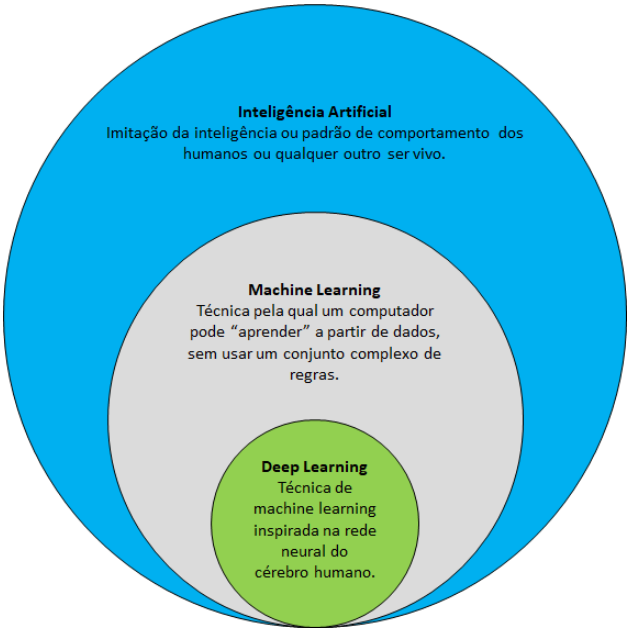
solucionar esse problema são: dispositivos de armazenamento local, sistemas de armazenamento em nuvem ou ambos. A solução escolhida depende dos requisitos de processamento desejados e dos mecanismos de processamento necessários para a análise dos dados. Tradicionalmente, usa-se armazenamento locais para os dados, porém, com popularização das nuvens, que deixaram de ser somente locais remotos de armazenamento e passaram a contar com ferramentas dedicadas de processamento de dados, tem havido uma migração pelo menos parcial para esse novo modelo (LOHR, 2012) (CUI, 2016) (O'MALLEY, 2014).

Por fim, uma vez que os dados estejam organizados e armazenados adequadamente, pode-se atuar efetivamente na análise, a fim de que o investimento em *Big Data* produza os retornos desejados, conferindo mais clareza com a análise visual dos diferentes conjuntos de dados e se aprofundando nos dados para fazer novas descobertas. A tendência é que sejam utilizados sistemas de inteligência artificial e aprendizado de máquina nesta etapa (LOHR, 2012) (CUI, 2016) (O'MALLEY, 2014).

2.3 Deep Learning

Deep learning (DL), termo em inglês que pode ser traduzido como aprendizado profundo, é uma das técnicas presentes na família do aprendizado de máquina (vide Figura 2) e se baseia em redes neurais artificiais com aprendizado de representação. O aprendizado pode ser supervisionado, semi-supervisionado, não supervisionado ou por reforço, dependendo da maneira como os dados são apresentados para o modelo a ser treinado (BENGIO, COURVILLE e VINCENT, 2013) (SCHMIDHUBER, 2015) (BENGIO, LECUN e HINTON, 2015).

Figura 2: Relação entre Inteligência Artificial, Machine Learning e Deep Learning.



Fonte: Adaptado de (SRIVASTAV, 2020)

Existem várias arquiteturas de DL, que são classificadas conforme a sua construção. Entre as mais conhecidas, pode-se citar as redes neurais profundas (*Deep Neural Networks* – DNNs), redes neurais recorrentes (*Recurrent Neural Networks* – RNNs) e redes neurais convolucionais (*Convolutional Neural Networks* – CNNs). Essas redes neurais já vêm sendo aplicadas nos campos de visão computacional, reconhecimento de fala, processamento de linguagem natural, tradução automática, bioinformática, *design* de drogas, análise de imagens médicas, inspeção de materiais e programas de jogos de tabuleiro. Em muitas dessas aplicações, as redes neurais são capazes de produzir resultados comparáveis e, em alguns casos, superiores ao desempenho de especialistas humanos (CIRESAN, MEIER e SCHMIDHUBER, 2012) (KRIZHEVSKY, SUTSKEVER e HINTON, 2012) (TECHCRUNCH, 2017) (HU, NIU, *et al.*, 2020).

As redes neurais artificiais (*Artificial Neural Networks* - ANNs), nome genérico que inclui todos os modelos supracitados, são inspiradas no processamento de informações e a maneira como elas são distribuídas em sistemas biológicos. Contudo, as ANNs têm várias diferenças em relação aos cérebros biológicos, pois tendem a ser estáticas, ao contrário do cérebro biológico, que é dinâmico, característica também conhecida como plasticidade neural (MARBLESTONE, WAYNE e KORDING) (OLSHAUSEN, 1996) (BENGIO, LEE, *et al.*, 2015).

A profundidade de uma ANN se refere à quantidade de camadas de neurônios artificiais entre a entrada e a saída da rede. Trabalhos anteriores mostraram que um

perceptron linear (o modelo mais simples de ANN) não pode ser um classificador universal, mas uma ANN com função de ativação não polinomial com uma camada oculta de largura ilimitada pode

2.3.1 Definição

DL é uma classe de algoritmos de aprendizado de máquina construídos com várias camadas a fim de extrair progressivamente características do dado apresentado à entrada. Em aplicações de processamento de imagem, por exemplo, as primeiras camadas identificam linhas e cores, enquanto as camadas mais adiante (mais profundas) podem identificar como as características extraídas pelas primeiras camadas compõem conceitos mais abstratos e relevantes para um ser humano, como dígitos, letras ou rostos (DENG e YU, 2014).

2.3.2 Visão Geral

A maioria dos modelos modernos de DL são baseados em ANNs, especificamente CNNs, embora também existam aqueles que utilizam fórmulas proposicionais ou variáveis latentes dentro das camadas de modelos generativos profundos, como os nós em redes de crenças profundas (*Deep Belief Networks* - DBFs) (BENGIO, 2009).

Em modelos de DL, cada nível (camada) aprende a transformar os dados de entrada em uma representação um pouco mais abstrata e composta. Em um aplicativo de reconhecimento de imagem, por exemplo, a entrada bruta pode ser uma matriz de *pixels*, de forma que a primeira camada abstrairia os *pixels* referentes às bordas, a segunda camada comporia e codificaria arranjos de bordas, a terceira camada codificaria elementos compostos, como nariz e olhos, e a quarta camada reconheceria que a imagem contém um rosto. É importante ressaltar que, durante o treinamento, um modelo de DL pode aprender por si só qual é a distribuição ideal para a alocação de recursos. No entanto, essa capacidade não elimina completamente a necessidade de ajuste manual (ajuste fino), o que geralmente implica na modificação da arquitetura da ANN e dos parâmetros de treinamento (BENGIO, COURVILLE e VINCENT, 2013) (LECUN, BENGIO e HINTON, 2015).

A profundidade de uma ANN se refere à quantidade de camadas ocultas responsáveis pela transformação dos dados. Mais especificamente, os modelos de DL têm uma profundidade associada aos caminhos de atribuição de crédito (*Credit*

Assignment Path - CAP), que são cadeias de conexões entre o dado bruto (camada de entrada) e o resultado (camada de saída). Em ANNs *feedforward*, nas quais os dados atravessam a rede num único sentido, partindo da camada de entrada até a camada de saída, a profundidade dos CAPs coincide com o número de camadas dessa rede. Por outro lado, em RNNs, que contam com realimentações entre as camadas (memória), a profundidade dos CAPs é, teoricamente, ilimitada (SCHMIDHUBER, 2015). Conceitualmente, não há um limiar universalmente aceito que além do qual o aprendizado de máquina se torne “profundo”, porém, a maioria dos pesquisadores concorda que qualquer ANN com CAP maior que 2 pode ser considerado um modelo de DL, uma vez que redes com estruturas desse nível de complexidade já demonstraram ser aproximadores universais no que diz respeito à capacidade de emular qualquer função (SHIGEKI, 2019).

As arquiteturas de DL podem ser construídas metodicamente camada por camada, de forma que se possa escolher quais parâmetros melhoram o desempenho do modelo no processo de extração de características (BENGIO, LAMBLIN, *et al.*, 2007).

Para tarefas de aprendizado supervisionado, os métodos de DL eliminam a obrigatoriedade de engenharia de recursos (pré-processamento), pois, ao contrário dos modelos de aprendizado de máquina convencionais, eles conseguem traduzir os dados em representações intermediárias compactas semelhantes às obtidas com o processo de análise dos componentes principais (Principal Components Analysis – PCA) e derivam as características dos dados entre as camadas de forma a remover redundâncias. Algoritmos de DL também podem ser aplicados a tarefas de aprendizado não supervisionado, o que é um benefício considerável, visto que os dados não rotulados são mais abundantes do que os rotulados. Esse tipo de aplicação, em geral, é utilizado como pré-análise de dados desconhecidos com a finalidade de entender a distribuição das informações contidas no banco de dados (SCHMIDHUBER, 2015) (BENGIO, COURVILLE e VINCENT, 2013) (HINTON, 2009).

2.3.3 Interpretações

DNNs são, geralmente, interpretadas em termos do teorema da aproximação universal ou inferência probabilística, que diz respeito à capacidade que as redes neurais *feedforward* com uma única camada oculta de tamanho finito têm para aproximar funções contínuas. Em 1989, a primeira prova foi publicada por George

Cybenko para funções de ativação sigmóide (CYBENKO, 1989) e foi generalizada para arquiteturas multicamadas em 1991 por Kurt Hornik (HORNIK, 1991). Trabalhos recentes também mostraram que a aproximação universal também é válida para funções de ativação não limitadas, como a unidade linear retificada (Rectified Linear Unit – ReLU) (SONODA e MURATA, 2017).

Embora tenha-se mencionado que o teorema da aproximação universal se refira a redes com camadas de tamanho finito (largura limitada), a quantidade de camadas (profundidade) pode aumentar. Lu et al. (LU, 2017) provaram que, se a largura de uma rede neural profunda com ativação ReLU for estritamente maior do que a dimensão de entrada, a rede pode se aproximar de qualquer função integrável de *Lebesgue*. Por outro lado, se a largura for menor ou igual à dimensão de entrada, a DNN não é um aproximador universal.

A interpretação probabilística deriva do campo do aprendizado de máquina e considera a não linearidade da função de ativação como uma função de distribuição cumulativa (MURPHY, 2012). A interpretação probabilística levou à introdução do conceito de *dropout* como regularizador nas redes neurais (HINTON, SRIVASTAVA, *et al.*, 2012) e foi popularizada em pesquisas como a de Bishop (BISHOP, 2017).

2.3.4 História

Algumas fontes indicam que Frank Rosenblatt desenvolveu e explorou todos os elementos básicos dos sistemas de DL atuais (TAPPERT, 2019), tendo-os descrito em seu livro "Princípios de Neurodinâmica: Perceptrons e a Teoria dos Mecanismos do Cérebro", publicado pelo Cornell Aeronautical Laboratory, Inc., *Cornell University* em 1962.

O primeiro algoritmo geral de aprendizagem funcional para perceptrons supervisionados, profundos, *feedforward* e multicamadas foi publicado por Alexey Ivakhnenko e Lapa em 1967 (IVAKHNENKO e LAPA, 1967). Um artigo de 1971 descreveu uma rede profunda com oito camadas treinadas pelo método de grupo de tratamento de dados (IVAKHNENKO, 1971). Outras arquiteturas de aprendizado profundo, especificamente aquelas construídas para visão computacional, começaram com o *Neocognitron* introduzido por Kunihiko Fukushima em 1980 (FUKUSHIMA, 1980).

O termo *Deep Learning* foi introduzido na comunidade de aprendizado de máquina por Rina Dechter em 1986 (DECHTER, 2016) e nas redes neurais artificiais

por Igor Aizenberg e colegas em 2000, no contexto dos neurônios de limiar booleano (IGOR AIZENBERG) (Co-evolving recurrent neurons learn deep memory POMDPs. Proc, USA).

Em 1989, Yann LeCun et al. aplicaram o algoritmo de retropropagação (*backward propagation*) padrão, que existia como o modo reverso de diferenciação automática desde 1970, a uma rede neural profunda com o objetivo de reconhecer códigos postais escritos à mão no correio. Durante a execução do algoritmo, o treinamento demorou 3 dias (LINNAINMAA, 1970) (GRIEWANK, 2012) (WERBOS, 1974) (WERBOS, 1982) (AL., 1989).

Em 1991, tais sistemas eram usados para reconhecer dígitos bidimensionais (2D) manuscritos, enquanto o reconhecimento de objetos tridimensionais (3D) era feito combinando imagens 2D com um modelo de objeto 3D feito à mão. Weng et al. (WENG, AHUJA e HUANG, 1992) sugeriram que um cérebro humano não usa um modelo de objeto 3-D monolítico e, em 1992, publicaram Cresceptron, um método para realizar o reconhecimento de objetos 3D em cenas desordenadas. Por usar imagens naturais diretamente, a Cresceptron deu início ao aprendizado visual de propósito geral para mundos 3D naturais. Cresceptron é uma cascata de camadas semelhante ao *Neocognitron*. Mas enquanto a *Neocognitron* exigia um programador humano para mesclar os recursos manualmente, o Cresceptron aprendeu um número aberto de recursos em cada camada sem supervisão, onde cada recurso é representado por um *kernel* de convolução. A Cresceptron segmentou cada objeto aprendido a partir de uma cena desordenada por meio de análises retrospectivas através da rede. A função *max pooling*, muito popular atualmente em DNNs, como as utilizadas com o conjunto de dados *ImageNet*, foi usado pela primeira vez no *Cresceptron* para reduzir a resolução dos filtros da rede neural com um *kernel* de dimensão 2x2 em cascata para melhor generalização (DE CARVALHO, FAIRHURST e BISSET, 1994) (HINTON, DAYAN, *et al.*, 1995) (S. HOCHREITER., 2015).

Em 1994, André de Carvalho, juntamente com Mike Fairhurst e David Bisset, publicou resultados experimentais de uma rede neural booleana multicamadas, também conhecida como rede neural sem peso, composta por um módulo de rede neural de extração de características de 3 camadas seguido por um módulo de rede neural de classificação multicamadas, que foram treinados de forma independente. Cada camada no módulo de extração obteve recursos com complexidade crescente em relação à camada anterior (HOCHREITER, BENGIO, *et al.*, 2001).

Em 1995, Brendan Frey demonstrou que era possível treinar em dois dias uma rede contendo seis camadas totalmente conectadas e várias centenas de unidades ocultas usando o algoritmo vigília-sono, desenvolvido em parceria com Peter Dayan e Hinton (BEHNKE, 2003). Muitos fatores contribuem para a velocidade lenta, incluindo o problema do desaparecimento de gradiente (*vanishing gradient*) analisado em 1991 por Sepp Hochreiter (MORGAN, BOURLARD, *et al.*, 1983) (ROBINSON, 1992).

Desde 1997, Sven Behnke estendeu a abordagem convolucional hierárquica *feedforward* na pirâmide de abstração neural por conexões laterais e retrógradas a fim de incorporar de forma flexível o contexto nas decisões e resolver iterativamente as ambiguidades locais (WAIBEL, HANAZAWA, *et al.*, 1989).

Modelos mais simples, que usam recursos específicos para tarefas, como filtros Gabor e máquinas de vetores de suporte (Support Vector Machines - SVMs), se popularizaram nos anos 1990 e 2000, devido ao custo computacional das ANNs.

Tanto o aprendizado superficial quanto o profundo de ANNs foram explorados por muitos anos para aplicações de reconhecimento de fala (BAKER, DENG, *et al.*, 2009), mas esses métodos nunca superaram a tecnologia do modelo de mistura gaussiana e do modelo de Markov Oculto (Gaussian Mixture Model and Hidden Markov Model - GMM-HMM) (BENGIO, 1991) (DENG, HASSANEIN e ELMASRY, 1994), que se baseiam em modelos generativos de fala treinados discriminativamente (DODDINGTON, PRZYBOCKI, *et al.*, 2000). As principais dificuldades foram analisadas, incluindo diminuição do gradiente e a estrutura de correlação temporal fraca em modelos preditivos neurais (HECK, KONIG, *et al.*, 2000) (ZOLTÁN TÜSKE, 2014). Dificuldades adicionais foram a falta de dados de treinamento e capacidade de computação limitada.

A maioria dos pesquisadores de reconhecimento de fala se afastou das redes neurais para buscar a modelagem generativa. Uma exceção foi a *SRI International* no final dos anos 1990. Financiado pela NSA e DARPA do governo dos EUA, o SRI estudou redes neurais profundas em reconhecimento de voz e alto-falante. A equipe de reconhecimento de alto-falante liderada por Larry Heck relatou sucesso significativo com redes neurais profundas no processamento de fala na avaliação de 1998 do Instituto Nacional de Padrões e Tecnologia de Reconhecimento de Falante (HOCHREITER e SCHMIDHUBER, 1997). A rede neural profunda SRI foi então implantada no *Nuance Verifier*, representando a primeira grande aplicação industrial de aprendizado profundo (GRAVES, ECK, *et al.*, 2003).

O princípio de elevar os recursos "brutos" em relação à otimização artesanal foi explorado pela primeira vez com sucesso na arquitetura do autoencoder profundo no espectrograma "bruto" ou recursos de banco de filtros lineares no final da década de 1990 (GRAVES, ECK, *et al.*, 2003), mostrando sua superioridade sobre o Mel-Recursos *cepstrais* que contêm estágios de transformação fixa a partir de espectrogramas. As características brutas da fala, formas de onda, posteriormente produziram excelentes resultados em larga escala (GRAVES, FERNÁNDEZ e GOMEZ, 2006).

Muitos aspectos do reconhecimento de fala foram assumidos por um método de aprendizado profundo denominado memória de curto e longo prazo (Long Short-Term Memory - LSTM), uma rede neural recorrente publicada por Hochreiter e Schmidhuber em 1997 (SANTIAGO FERNANDEZ, 2007). As RNNs com LSTM evitam o problema do gradiente de desaparecimento e podem aprender tarefas de "Aprendizado Muito Profundo" que requerem memórias de eventos que aconteceram milhares de passos de tempo discretos antes, o que é importante para a fala. Em 2003, o LSTM começou a se tornar competitivo com os reconhecedores de fala tradicionais em certas tarefas (SAK, SENIOR, *et al.*, 2015). Mais tarde, foi combinada com a classificação temporal conexionista (Connectionist Temporal Classification - CTC) (HINTON, 2007) em pilhas de RNNs LSTM (HINTON, OSINDERO e TEH, 2015). Em 2015, o reconhecimento de voz do *Google* supostamente experimentou um salto dramático de desempenho de 49% por meio do LSTM treinado pelo CTC, que foi disponibilizado por meio da pesquisa por voz do *Google* (BENGIO, 2012).

Em 2006, publicações de Geoff Hinton, Ruslan Salakhutdinov, Osindero e Teh (G. E. HINTON., 2007) (HINTON, DENG, *et al.*, 2012) (DENG, HINTON e KINGSBURY, 2017) mostraram como uma rede neural *feedforward* de muitas camadas poderia ser efetivamente pré-treinada uma camada por vez, tratando cada camada individualmente, como uma máquina de Boltzmann restrita não supervisionada e, em seguida, ajustando-a usando retropropagação supervisionada (DENG, LI, *et al.*, 2013). Os artigos referiam-se ao aprendizado para redes de crenças profundas.

O aprendizado profundo faz parte de sistemas de última geração em várias disciplinas, particularmente em visão computacional e reconhecimento automático de fala (Automatic Speech Recognition - ASR). Os resultados em conjuntos de avaliação comumente usados, como TIMIT (ASR) e MNIST (classificação de imagem), bem

como uma variedade de tarefas de reconhecimento de fala de grande vocabulário, têm melhorado continuamente (SINGH, SAHA e SAHIDULLAH, 2021) (SAK, SENIOR e BEAUFAYS, 2014) (LI e WU, 2014) (ZEN e SAK, 2015) (DENG, ABDEL-HAMID e YU, 2013).

O impacto do aprendizado profundo na indústria começou no início dos anos 2000, quando as CNNs já processavam cerca de 10% a 20% de todos os cheques emitidos nos Estados Unidos. As aplicações industriais de aprendizado profundo para reconhecimento de fala em grande escala começaram por volta de 2010 (D. YU, 2011).

O *Workshop* NIPS de 2009 sobre Aprendizado Profundo para Reconhecimento de Fala foi motivado pelas limitações dos modelos generativos profundos de fala e pela possibilidade de que, dado um *hardware* mais capaz e conjuntos de dados em grande escala, as DNNs pudessem se tornar práticas. Acreditava-se que o pré-treinamento de DNNs usando modelos generativos de DBNs superariam as principais dificuldades das redes neurais. No entanto, foi descoberto que substituir o pré-treinamento por grandes quantidades de dados de treinamento para retropropagação direta ao usar DNNs com grandes camadas de saída dependentes do contexto produzia taxas de erro drasticamente menores do que o modelo GMM-HMM e também de sistemas baseados em modelos generativos mais avançados. A natureza dos erros de reconhecimento produzidos pelos dois tipos de sistemas era caracteristicamente diferente, oferecendo *insights* técnicos sobre como integrar o aprendizado profundo no sistema de decodificação de fala em tempo de execução altamente eficiente existente implantado por todos os principais reconhecimentos de sistemas de fala. A análise por volta de 2009-2010, contrastando o GMM (e outros modelos de fala generativos) com modelos DNN, estimulou o investimento industrial inicial em aprendizagem profunda para reconhecimento de fala, eventualmente levando ao uso generalizado e dominante nessa indústria. Essa análise foi feita com desempenho comparável (menos de 1,5% na taxa de erro) entre DNNs discriminativos e modelos generativos (DENG, HINTON e KINGSBURY, 2017) (YU e DENG, 2014) (LI, 2014).

Em 2010, os pesquisadores estenderam o aprendizado profundo do TIMIT para o reconhecimento de voz de grande vocabulário, adotando grandes camadas de saída do DNN com base em estados HMM dependentes do contexto construídos por árvores de decisão (YU e DENG, 2010).

Os avanços no *hardware* geraram um interesse renovado no aprendizado profundo. Em 2009, a Nvidia estava envolvida no que foi chamado de “*big bang*” do aprendizado profundo, “já que as redes neurais de aprendizado profundo eram treinadas com unidades de processamento gráfico (GPUs) da Nvidia.” Naquele ano, Andrew Ng determinou que as GPUs poderiam aumentar a velocidade dos sistemas de aprendizado profundo em cerca de 100 vezes. Em particular, as GPUs são adequadas para os cálculos de matriz / vetor envolvidos no aprendizado de máquina (SZE, CHEN, *et al.*, 2017). As GPUs aceleram os algoritmos de treinamento em ordens de magnitude, reduzindo os tempos de execução de semanas para dias, além disso, *hardware* especializado e otimizações de algoritmo podem ser usados para processamento eficiente de modelos de aprendizado profundo (NATIONAL CENTER FOR ADVANCING TRANSLATIONAL SCIENCES, 2014).

2.3.5 Redes Neurais

Em 2012, uma equipe liderada por George E. Dahl venceu o "*Merck Molecular Activity Challenge*" usando redes neurais profundas multitarefas para prever o alvo biomolecular de um medicamento (NATIONAL CENTER FOR ADVANCING TRANSLATIONAL SCIENCES, 2014). Em 2014, o grupo de *Hochreiter* usou o aprendizado profundo para detectar efeitos tóxicos e fora do alvo de produtos químicos ambientais em nutrientes, produtos domésticos e drogas e venceu o "Desafio de Dados Tox21" do NIH, FDA e NCATS. (CIRESAN, MEIER, *et al.*, 2021) (CIRESAN, GIUSTI, *et al.*, 2012) (CIRESAN, GIUSTI, *et al.*, 2013)

Impactos adicionais significativos na imagem ou reconhecimento de objeto foram sentidos de 2011 a 2012. Embora CNNs treinadas com *backpropagation* já existissem por décadas, inclusive com implementações de GPU, maiores evoluções eram necessárias para progredir na visão computacional. Em 2011, essa abordagem alcançou, pela primeira vez, um desempenho sobre-humano em um concurso de reconhecimento de padrões visuais. Até 2011, as CNNs não desempenhavam um papel importante nas conferências de visão computacional, mas, em junho de 2012, um artigo de Ciresan et al. (CIRESAN, GIUSTI, *et al.*, 2012) mostrou como CNNs na GPU podem melhorar drasticamente muitos registros de *benchmark* de visão. Em outubro de 2012, um sistema semelhante por Krizhevsky et al. (KRIZHEVSKY, SUTSKEVER e HINTON, 2012) venceu a competição em grande escala da *ImageNet* por uma margem significativa sobre os métodos superficiais de aprendizado de

máquina. Em novembro de 2012, o sistema de Ciresan et al. (CIRESAN, MEIER e SCHMIDHUBER, 2012) também ganhou o concurso ICPR na análise de grandes imagens médicas para detecção de câncer e, no ano seguinte, também o Grande Desafio MICCAI sobre o mesmo tópico. Em 2013 e 2014, a taxa de erro na tarefa *ImageNet* usando aprendizado profundo foi reduzida ainda mais, seguindo uma tendência semelhante no reconhecimento de fala em larga escala (KIROS, SALAKHUTDINOV e ZEMEL, 2014).

A classificação de imagens foi então estendida para a tarefa mais desafiadora de gerar descrições (legendas) para imagens, frequentemente como uma combinação de CNNs e LSTMs (ZHONG, LIU e LIU, 2011) (SILVER, HUANG, *et al.*, 2016).

Alguns pesquisadores afirmam que a vitória do *ImageNet* em outubro de 2012 ancorou o início de uma "revolução do aprendizado profundo" que transformou a indústria de IA (SZEGEDY, TOSHEV e ERHAN, 2013).

Em março de 2019, Yoshua Bengio, Geoffrey Hinton e Yann LeCun receberam o Prêmio Turing por inovações conceituais e de engenharia que tornaram as redes neurais profundas um componente crítico da computação.

2.3.5.1 Redes Neurais Artificiais

As ANNs são sistemas de computação inspirados nas redes neurais biológicas que constituem o cérebro dos animais. Esses sistemas aprendem (melhoram progressivamente sua capacidade) a realizar tarefas considerando exemplos, geralmente sem programação específica para tarefas. Por exemplo, no reconhecimento de imagem, eles podem aprender a identificar imagens que contêm gatos analisando imagens de exemplo que foram marcadas manualmente como "com gato" ou "sem gato" e usando os resultados analíticos para identificar gatos em outras imagens. Eles descobriram que a maior parte do uso em aplicativos é difícil de expressar com um algoritmo de computador tradicional usando programação baseada em regras.

Uma ANN se baseia em uma coleção de unidades conectadas chamadas neurônios artificiais (análogos aos neurônios biológicos em um cérebro biológico). Cada conexão (sinapse) entre neurônios pode transmitir um sinal para outro neurônio. O neurônio receptor (pós-sináptico) pode processar o sinal e sinalizar os neurônios a jusante conectados a ele. Os neurônios têm estados representados por números

reais, normalmente entre 0 e 1. Os neurônios e as sinapses também podem ter um peso que varia à medida que o aprendizado avança, o que pode aumentar ou diminuir a força do sinal que envia para adiante.

Normalmente, os neurônios são organizados em camadas. Diferentes camadas podem realizar diferentes tipos de transformações em suas entradas. Os sinais viajam da camada de entrada até a camada de saída depois de atravessar um conjunto de camadas ocultas.

O objetivo original da abordagem da rede neural era resolver problemas da mesma forma que um cérebro humano faria. Com o tempo, a atenção se concentrou em combinar habilidades mentais específicas, levando a desvios da biologia, como retropropagação, ou passar informações na direção reversa e ajustar a rede para refletir essas informações.

As redes neurais têm sido usadas em uma variedade de tarefas, incluindo visão computacional, reconhecimento de fala, tradução automática, filtragem de redes sociais, jogos de tabuleiro e videogames e diagnóstico médico.

A partir de 2017, as redes neurais normalmente têm alguns milhares a alguns milhões de unidades e milhões de conexões. Apesar deste número ser várias ordens de magnitude menor do que o número de neurônios em um cérebro humano, essas redes podem realizar muitas tarefas em um nível além do dos humanos (ROLNICK e TEGMARK, 2018).

2.3.5.2 Redes Neurais Profundas

Uma rede neural profunda (DNN) é uma ANN com várias camadas entre as camadas de entrada e saída. Existem diferentes tipos de redes neurais, mas sempre consistem nos mesmos componentes: neurônios, sinapses, pesos, vieses e funções. Esses componentes funcionam de forma semelhante ao cérebro humano e podem ser treinados como qualquer outro algoritmo de aprendizado de máquina (HOF, 2018) (SCHMIDHUBER, 2015).

Por exemplo, uma DNN treinada para reconhecer raças de cães examinará a imagem fornecida e calculará a probabilidade de o cão na imagem ser uma determinada raça. O usuário pode revisar os resultados e selecionar quais probabilidades a rede deve exibir (acima de um certo limite, etc.) e retornar o rótulo proposto. Cada manipulação matemática como tal é considerada uma camada, e DNN complexos têm muitas camadas, daí o nome de redes "profundas".

As DNNs podem modelar relacionamentos não lineares complexos, uma vez que suas arquiteturas geram modelos composicionais onde o objeto é expresso como uma composição de elementos extraídos nas camadas primitivas (GERS e SCHMIDHUBER, 2001). As camadas extras permitem a composição de recursos de camadas inferiores, potencialmente modelando dados complexos com menos unidades do que uma rede rasa de desempenho semelhante. Por exemplo, foi provado que polinômios multivariados esparsos são exponencialmente mais fáceis de aproximar com DNNs do que com redes rasas (SUTSKEVER, VINYALS e LE, 2014).

As arquiteturas profundas incluem muitas variantes de algumas abordagens básicas. Cada arquitetura obteve sucesso em domínios específicos. Nem sempre é possível comparar o desempenho de várias arquiteturas, a menos que tenham sido avaliadas nos mesmos conjuntos de dados.

DNNs são normalmente redes *feedforward* nas quais os dados fluem da camada de entrada para a camada de saída sem *loopback*. A princípio, a DNN cria um mapa de neurônios virtuais e atribui valores numéricos aleatórios, ou pesos, às conexões entre eles. Os pesos e entradas são multiplicados e retornam uma saída entre 0 e 1. Se a rede não reconhecesse com precisão um padrão particular, um algoritmo ajustaria os pesos. Dessa forma, o algoritmo pode tornar certos parâmetros mais influentes, até determinar a manipulação matemática correta para processar totalmente os dados (SOCHER e MANNING, 2014).

Redes neurais recorrentes (RNNs), nas quais os dados podem fluir em qualquer direção, ao contrário das DNNs convencionais, nas quais não há retorno entre as camadas, são usadas para aplicações como modelagem de linguagem e o modelo LSTM é particularmente eficaz para esse uso. (DAHL, SAINATH e HINTON, 2013)

Redes neurais convolucionais (CNNs) são usadas na visão computacional. As CNNs também foram aplicadas à modelagem acústica para reconhecimento automático de fala (ASR) (COURSERA).

2.3.5.3 Desafios

Tal como acontece com ANNs, muitos problemas podem surgir durante o treinamento de DNNs, dentre os quais, pode-se citar o *overfitting* e o tempo de computação. DNNs são propensos a *overfitting* devido às camadas adicionadas de abstração, que permitem modelar dependências raras nos dados de treinamento.

Alternativamente, a regularização por meio de *dropout* omite, aleatoriamente, unidades das camadas ocultas durante o treinamento, de forma a ajudar a excluir dependências raras. Finalmente, os dados podem ser aumentados por meio de métodos como recorte e rotação (*data augmentation*), de modo que conjuntos de treinamento menores possam ser aumentados em tamanho para reduzir as chances de *overfitting* (VIEBKE, MEMETI, *et al.*, 2019).

As DNNs devem considerar muitos parâmetros de treinamento, como o tamanho (número de camadas e número de unidades por camada), a taxa de aprendizado (*learning rate*) e os pesos iniciais (geralmente, aleatórios). A varredura através do espaço de parâmetros para parâmetros ideais pode não ser viável devido ao custo em tempo e recursos computacionais. Vários truques, como *batching* (computar o gradiente em vários exemplos de treinamento de uma vez em vez de exemplos individuais) aceleram o cálculo. Grandes capacidades de processamento de arquiteturas de muitos núcleos (como GPUs) produziram acelerações significativas no treinamento, devido à adequação de tais arquiteturas de processamento para cálculos matriciais e vetoriais (TING QIN, 2005).

Como alternativa, os engenheiros podem procurar outros tipos de redes neurais com algoritmos de treinamento mais diretos e convergentes. O controlador de articulação do modelo cerebelar (*Cerebellar Model Articulation Controller* – CMAC) é um desses tipos de rede neural e não requer taxas de aprendizagem ou pesos iniciais aleatórios. O processo de treinamento pode ser garantido para convergir em uma etapa com um novo lote de dados, e a complexidade computacional do algoritmo de treinamento é linear com relação ao número de neurônios envolvidos (KOBIELUS, 2019).

2.3.6 Aplicações

2.3.6.1 Reconhecimento automático de voz

O reconhecimento automático de voz em grande escala é o primeiro e mais convincente caso de sucesso de aprendizagem profunda. RNNs LSTM podem aprender tarefas de que envolvem intervalos de vários segundos contendo eventos de fala separados por milhares de intervalos de tempo discretos, onde um intervalo de tempo corresponde a cerca de 10 ms (DAHL, SAINATH e HINTON, 2013).

O sucesso inicial no reconhecimento de fala foi baseado em tarefas de reconhecimento em pequena escala baseadas no TIMIT. O conjunto de dados contém

630 falantes de oito dialetos principais do inglês americano, onde cada falante lê 10 sentenças. Seu pequeno tamanho permite que muitas configurações sejam experimentadas. Mais importante, a tarefa TIMIT diz respeito ao reconhecimento de sequência de telefone, que, ao contrário do reconhecimento de sequência de palavras, permite modelos de linguagem bigrama de telefone fracos. Isso permite que a força dos aspectos de modelagem acústica do reconhecimento de voz seja analisada mais facilmente. As taxas de erro listadas na Tabela 1, incluindo esses resultados iniciais e medidas como taxas de erro percentual, foram obtidas desde 1991 (ROBINSON, 1991).

Tabela 1: Taxas de erros

Método	**Taxa de Erro Percentual**
RNN inicializado aleatoriamente	26,1
Trifone Bayesiano GMM-HMM	25,6
Modelo de trajetória oculta (gerador)	24,8
DNN de monofone inicializado aleatoriamente	23,4
Monofone DBN-DNN	22,4
Triphone GMM-HMM com treinamento BMMI	21,7
Monofone DBN-DNN no fbank	20,7
DNN convolucional	20,0
DNN convolucional w. Pooling heterogêneo	18,7
Conjunto DNN / CNN / RNN	18,3
LSTM bidirecional	17,8
Rede Hierárquica Convolucional Deep Maxout	16,5

Fonte: Vaz Gabriel (2023)

A estreia das DNNs para reconhecimento de fala no final da década de 1990 e reconhecimento de fala por volta de 2009-2011 e do LSTM por volta de 2003-2007, acelerou o progresso em oito áreas principais (BENGIO, COURVILLE e VINCENT, 2013):

- Expansão / redução e treinamento e decodificação de DNN acelerado;
- Treinamento discriminativo de sequência;
- Processamento de recursos por modelos profundos com conhecimento sólido dos mecanismos subjacentes;
- Adaptação de DNNs e modelos de profundidade relacionados;
- Multi-tarefa e aprendizagem de transferência por DNNs e modelos de profundidade relacionados;
- CNNs e como projetá-los para melhor explorar o conhecimento do domínio da fala;
- RNN e suas ricas variantes LSTM;
- Outros tipos de modelos profundos, incluindo modelos baseados em tensores e modelos profundos generativos / discriminativos integrados;

Todos os principais sistemas de reconhecimento de voz comerciais (por exemplo, Microsoft Cortana, Xbox, Skype Translator, Amazon Alexa, Google Now, Apple Siri, Baidu e pesquisa de voz iFlyTek e uma variedade de produtos de fala Nuance, etc.) são baseados em aprendizagem profunda (CIREşAN, MEIER, *et al.*, 2011).

2.3.6.2 Reconhecimento de imagem

Um conjunto de avaliação comum para classificação de imagens é o conjunto de dados do banco de dados MNIST. MNIST é composto de dígitos manuscritos e inclui 60.000 exemplos de treinamento e 10.000 exemplos de teste. Assim como o TIMIT, seu tamanho pequeno permite que os usuários testem várias configurações. Uma lista abrangente de resultados neste conjunto está disponível (SMITH e LEYMARIE, 2017).

O reconhecimento de imagem baseado em aprendizado profundo tornou-se "sobre-humano", produzindo resultados mais precisos do que competidores humanos. Isso ocorreu pela primeira vez em 2011, em reconhecimento de sinais de trânsito, e em 2014, com o reconhecimento de rostos humanos (*Surpassing Human Level Face Recognition*) (ARCAS, 2017).

Veículos treinados em aprendizagem profunda agora interpretam visualizações de câmera de 360 °. Outro exemplo é o *Facial Dismorphology Novel Analysis*, usado

para analisar casos de malformação humana conectados a um grande banco de dados de síndromes genéticas (GOLDBERG e LEVY, 2014).

2.3.6.3 Processamento de arte visual

Intimamente relacionado ao progresso feito no reconhecimento de imagens, está a aplicação crescente de técnicas de aprendizado profundo em várias tarefas de artes visuais. DNNs provaram ser capazes, por exemplo: (SOCHER e MANNING, 2014)

a) Identificar o período de estilo de uma dada pintura;
b) Transferência de estilo neural - capturar o estilo de uma determinada obra de arte e aplicá-lo de uma maneira visualmente agradável a uma fotografia ou vídeo arbitrário;
c) Geração de imagens impressionantes com base em campos de entrada visual aleatórios.

2.3.6.4 Processamento de linguagem natural

As redes neurais têm sido usadas para implementar modelos de linguagem desde o início dos anos 2000 (GILLICK, BRUNK, *et al.*, 2015) e o LSTM ajudou a melhorar a tradução automática e a modelagem de linguagem (MIKOLOV e AL., 2010).

Outras técnicas importantes neste campo são a amostragem negativa e a incorporação de palavras. A incorporação de palavras, como *word2vec*, pode ser considerada uma camada representacional em uma arquitetura de aprendizado profundo que transforma uma palavra atômica em uma representação posicional da palavra em relação a outras palavras no conjunto de dados; a posição é representada como um ponto em um espaço vetorial. Usar a incorporação de palavras como uma camada de entrada RNN permite que a rede analise sentenças e frases usando uma gramática de vetor de composição eficaz. Uma gramática de vetor de composição pode ser considerada como uma gramática livre de contexto probabilística implementada por um RNN. Codificadores automáticos recursivos construídos sobre *embeddings* de palavras podem avaliar a semelhança de frases e detectar paráfrases. Arquiteturas neurais profundas fornecem os melhores resultados para análise de constituintes, análise de sentimento, recuperação de informação, compreensão da linguagem falada, tradução automática, vinculação de entidade contextual,

reconhecimento de estilo de escrita, classificação de texto e outros (RICHARD SOCHER, 2013).

Desenvolvimentos recentes generalizam a incorporação de palavras para a incorporação de frases.

O *Google Translate* usa uma grande rede de memória de longo prazo (LSTM) ponta a ponta. A tradução automática do *Google Neural* usa um método de tradução automática baseado em exemplos em que o sistema "aprende com milhões de exemplos". Ele traduz "frases inteiras de uma vez, em vez de partes. O *Google Translate* oferece suporte a mais de cem idiomas e a rede codifica a semântica da frase ao invés de simplesmente memorizar traduções frase a frase, usando o inglês como um intermediário entre a maioria dos pares de idiomas (WU, 2016) (THE GLOBE AND MAIL, 2015).

2.3.6.5 Descoberta de drogas e toxicologia

Uma grande porcentagem de medicamentos candidatos não consegue obter a aprovação regulatória. Essas falhas são causadas por eficácia insuficiente (efeito no alvo), interações indesejadas (efeitos fora do alvo) ou efeitos tóxicos imprevistos (GARLING, 2015). A pesquisa explorou o uso do aprendizado profundo para prever os alvos biomoleculares, fora dos alvos e os efeitos tóxicos de produtos químicos ambientais em nutrientes, produtos domésticos e drogas (CIRESAN, MEIER, *et al.*, 2011).

AtomNet é um sistema de aprendizado profundo para o projeto racional de drogas baseado em estrutura. O AtomNet foi usado para prever novas biomoléculas candidatas a alvos de doenças, como o vírus Ebola e a esclerose múltipla (ZHAVORONKOV, 2019) (GREGORY, 2020).

Em 2017, redes neurais de gráfico foram usadas pela primeira vez para prever várias propriedades de moléculas em um grande conjunto de dados de toxicologia. Em 2019, redes neurais generativas foram usadas para produzir moléculas que foram validadas experimentalmente em ratos (VAN DEN OORD, DIELEMAN e SCHRAUWEN, 2013) (FENG, ZHANG, *et al.*, 2019).

2.3.6.6 Sistemas de recomendação

Os sistemas de recomendação têm usado o aprendizado profundo para extrair recursos significativos para um modelo de fator latente para música baseada em

conteúdo e recomendações de periódicos (CHICCO, SADOWSKI e BALDI, 2014) (SATHYANARAYANA, 2016). O aprendizado profundo de múltiplas visualizações foi aplicado para aprender as preferências do usuário em vários domínios (CHOI, SCHUETZ, *et al.*, 2016) e seus modelos usam uma abordagem híbrida colaborativa e baseada em conteúdo e aprimoram as recomendações em várias tarefas.

2.3.6.7 Bioinformática

Uma ANN autoencoder foi usada em bioinformática, para prever anotações de ontologia gênica e relações de função gênica (LITJENS, KOOI, *et al.*, 2017,). Em informática médica, o aprendizado profundo foi usado para prever a qualidade do sono com base em dados de *wearables* e previsões de complicações de saúde a partir de dados de registros eletrônicos de saúde (FORSLID, WIESLANDER, *et al.*, 2017) (DE, MAITY, *et al.*, 2017).

2.3.6.8 Análise de imagem médica

Foi demonstrado que o aprendizado profundo produz resultados competitivos em aplicações médicas, como classificação de células cancerosas, detecção de lesões, segmentação de órgãos e aprimoramento de imagem (DeOldify: Colorizing and Restoring Old Images and Videos with Deep Learning, 2018).

2.3.6.9 Publicidade móvel

Encontrar o público móvel apropriado para a publicidade móvel é sempre desafiador, uma vez que muitos pontos de dados devem ser considerados e analisados antes que um segmento-alvo possa ser criado e usado na veiculação de anúncios por qualquer servidor de anúncios (KLEANTHOUS e CHATZIS, 2020). O aprendizado profundo tem sido usado para interpretar grandes conjuntos de dados de publicidade de várias dimensões. Muitos pontos de dados são coletados durante o ciclo de solicitação / veiculação / clique de publicidade na Internet. Essas informações podem formar a base do aprendizado de máquina para melhorar a seleção de anúncios.

2.3.6.10 Restauração de imagem

O aprendizado profundo foi aplicado com sucesso a problemas inversos, como *denoising*, super-resolução, pintura interna e colorização de filme (CZECH, 2018).

Essas aplicações incluem métodos de aprendizagem, como "Campos de redução para restauração de imagem eficaz" (LABORATORY, 2018), que treina em um conjunto de dados de imagem, e *Deep Image Prior*, que treina na imagem que precisa ser restaurada.

2.3.6.11 Detecção de fraude financeira

O aprendizado profundo está sendo aplicado com sucesso à detecção de fraude financeira, detecção de evasão fiscal e combate à lavagem de dinheiro. (ELMAN, 1988)

2.3.6.12 Militar

O Departamento de Defesa dos Estados Unidos aplicou o aprendizado profundo para treinar robôs em novas tarefas por meio da observação. (SHRAGER e JOHNSON, 1996)

2.3.7 Relação com o desenvolvimento cognitivo e cerebral humano

O aprendizado profundo está intimamente relacionado a uma classe de teorias do desenvolvimento do cérebro (especificamente, desenvolvimento neocortical) propostas por neurocientistas cognitivos no início da década de 1990. Essas teorias de desenvolvimento foram instanciadas em modelos computacionais, tornando-os predecessores dos sistemas de aprendizado profundo. Esses modelos de desenvolvimento compartilham a propriedade de que várias dinâmicas de aprendizagem propostas no cérebro (por exemplo, uma onda de fator de crescimento do nervo) apoiam a auto-organização de alguma forma análoga às redes neurais utilizadas em modelos de aprendizagem profunda. Como o neocórtex, as redes neurais empregam uma hierarquia de filtros em camadas em que cada camada considera as informações de uma camada anterior (ou do ambiente operacional) e, em seguida, passa sua saída (e possivelmente a entrada original) para outras camadas. Este processo produz uma pilha auto-organizada de transdutores, bem ajustados ao seu ambiente operacional. Uma descrição de 1995 afirmou: "... o cérebro do bebê parece se organizar sob a influência de ondas dos chamados fatores tróficos ... diferentes regiões do cérebro se conectam sequencialmente, com uma camada de tecido amadurecendo antes da outra e assim até que todo o cérebro esteja maduro" (TESTOLIN e ZORZI, 2016).

Uma variedade de abordagens tem sido usada para investigar a plausibilidade dos modelos de aprendizagem profunda de uma perspectiva neurobiológica. Por outro lado, várias variantes do algoritmo de retropropagação foram propostas a fim de aumentar seu realismo de processamento (TESTOLIN, STOIANOV e ZORZI, 2017) (BUESING, BILL, *et al.*, 2011). Outros pesquisadores argumentaram que as formas não supervisionadas de aprendizagem profunda, como aquelas baseadas em modelos gerativos hierárquicos e redes de crenças profundas, podem estar mais próximas da realidade biológica (MOREL, SINGH e LEVY, 2010) (CASH e YUSTE, 1999). A esse respeito, os modelos de rede neural generativa foram relacionados a evidências neurobiológicas sobre o processamento baseado em amostragem no córtex cerebral (OLSHAUSEN e FIELD, 2004).

Embora uma comparação sistemática entre a organização do cérebro humano e a codificação neuronal em redes profundas ainda não tenha sido estabelecida, várias analogias foram relatadas. Por exemplo, os cálculos realizados por unidades de aprendizagem profunda podem ser semelhantes aos de neurônios reais (YAMINS e DICARLO, 2016) (ZORZI e TESTOLIN, 2018) e populações neurais (GÜÇLÜ e VAN GERVEN, 2015). Da mesma forma, as representações desenvolvidas por modelos de aprendizagem profunda são semelhantes àquelas medidas no sistema visual primata (METZ, 2014), tanto nos níveis de unidade única (GIBNEY, 2016) e na população (SILVER, HUANG, *et al.*, 2016).

2.3.8 Atividade comercial

O laboratório de IA do Facebook realiza tarefas como marcar automaticamente as fotos carregadas com os nomes das pessoas que estão nelas (MIT TECHNOLOGY REVIEW, 2016).

A *DeepMind Technologies* do Google desenvolveu um sistema capaz de aprender a jogar vídeo games Atari usando apenas *pixels* como entrada de dados. Em 2015, eles demonstraram seu sistema AlphaGo, que aprendeu o jogo Go bem o suficiente para vencer um jogador profissional de Go (BURNS, 2015) (METZ, 2019) (BRADLEY KNOX e STONE, 2008).

Em 2015, Blippar demonstrou um aplicativo de realidade aumentada móvel que usa aprendizado profundo para reconhecer objetos em tempo real (MCCANEY, 2018).

Em 2017, Covariant.ai foi lançado, com foco na integração do aprendizado profundo em fábricas (MARCUS, 2018).

A partir de 2008, (KNIGHT, 2017) pesquisadores da Universidade do Texas em Austin (UT) desenvolveram uma estrutura de aprendizado de máquina chamada Treinamento manual de um agente via reforço avaliativo, ou TAMER, que propôs novos métodos para robôs ou programas de computador aprenderem a executar tarefas interagindo com um instrutor humano (SHRAGER e JOHNSON, 1996). Desenvolvido pela primeira vez como TAMER, um novo algoritmo denominado Deep TAMER foi posteriormente introduzido em 2018 durante uma colaboração entre o Laboratório de Pesquisa do Exército dos EUA (ARL) e pesquisadores da UT. O Deep TAMER usou o aprendizado profundo para fornecer a um robô a habilidade de aprender novas tarefas por meio da observação (SHRAGER e JOHNSON, 1996). Usando Deep TAMER, um robô aprendeu uma tarefa com um treinador humano, assistindo a *streams* de vídeo ou observando um humano realizar uma tarefa pessoalmente. Mais tarde, o robô praticou a tarefa com a ajuda de algum *coaching* do treinador, que forneceu *feedback* como “bom trabalho” e “mau trabalho”. (MARCUS, 2017)

3 O USO DE INTELIGÊNCIA ARTIFICIAL E *DEEP LEARNING* NA ÁREA DA SAÚDE E BIOMÉDICA

Neste capítulo, serão apresentados dezesseis trabalhos nacionais e internacionais com a finalidade de contextualizar a temática abordada nesta pesquisa e expor ao leitor a importância e relevância do tema. O critério de semelhança de temáticas, assim como a fator de impacto e ano de publicação, foi utilizado para selecionar os dezesseis trabalhos, de forma que eles servem como base e referenciamento comparativo para esta dissertação, cuja metodologia e diferencial serão abordados nos capítulos seguintes.

3.1 Machine learning of serum metabolic patterns encodes early-stage lung adenocarcinoma

A detecção precoce do câncer aumenta muito as chances de sucesso do tratamento, mas diagnósticos para alguns tumores, incluindo adenocarcinoma pulmonar (AP), são limitados. Um ideal diagnóstico em estágio inicial de AP para uso clínico em larga escala deve abordar a detecção rápida, ser um procedimento pouco invasivo e ter alto desempenho. Aqui, conduzimos o aprendizado de máquina do metabolismo sérico padrões para detectar AP em estágio inicial. Foram extraídos padrões metabólicos diretos pela espectrometria otimizada de massa de dessorção/ionização a laser assistida por partículas férricas em 1 s usando apenas 50 nL de soro. Foi definida uma faixa metabólica de 100–400 Da com recursos de 143 m/z. Foi diagnosticada AP em estágio inicial com sensibilidade ~ 70–90% e especificidade ~ 90–93% por meio do esparso aprendizado de máquina de regressão de padrões. Identificamos um painel de biomarcadores de sete metabólitos e vias relevantes para distinguir AP em estágio inicial de controles ($p < 0,05$). A abordagem avança o projeto de análise metabólica para a detecção precoce do câncer e é promissor como um teste eficiente para implantação de baixo custo para clínicas (HUANG, WANG, *et al.*, 2020).

3.2 An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning

O aprendizado profundo para patologia digital é dificultado pela resolução espacial extremamente alta de imagens de slides inteiros (ISIs). A maioria dos estudos

empregou métodos baseados em adesivos, que muitas vezes exigem anotação detalhada de patches de imagem. Isso normalmente envolve uma mão de obra considerável para fazer os contornos nos ISIs. Para aliviar o peso desse trabalho e obter benefícios com a ampliação do treinamento com vários ISIs, desenvolvemos um método para treinar redes neurais em ISIs inteiros usando apenas diagnósticos em nível de slide. Nosso método aproveita o mecanismo de memória unificada para superar a restrição de memória dos aceleradores de computação. Experimentos conduzido em um conjunto de dados de 9.662 ISIs de câncer de pulmão revela que o método proposto atinge áreas sob a curva característica de operação do receptor de 0,9594 e 0,9414 para classificação de adenocarcinoma e carcinoma de células escamosas no conjunto de teste, respectivamente. Além disso, o método demonstra maior desempenho de classificação do que múltiplas instâncias aprendizado, bem como fortes resultados de localização para pequenas lesões por meio do mapeamento da ativação de classes (CHEN, CHEN, *et al.*, 2021).

3.3 Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography

Pacientes com câncer têm maior risco de mortalidade por doença cardiovascular (DCV) do que a população geral população. A tomografia computadorizada de baixa dose (TCBD) para rastreamento de câncer de pulmão oferece uma oportunidade para estimativa simultânea do risco de DCV em pacientes de risco. O modelo de aprendizado profundo de previsão de risco de DCV, treinado com 30.286 TCBDs do National Lung Cancer Screening Trial, atinge uma área sob a curva (Area Under the Curve - AUC) de 0,871 em um conjunto de teste separado de 2.085 indivíduos e identifica pacientes com alto risco de mortalidade por DCV (AUC de 0,768). O modelo foi validado contra marcadores baseados em tomografias cardíacas controlada por eletrocardiograma de um conjunto de dados independente de 335 indivíduos. O trabalho mostra que, em pacientes de alto risco, aprendizado profundo pode converter TCBD para triagem de câncer de pulmão em uma ferramenta quantitativa de triagem dupla para estimativa de risco de DCV (CHAO, SHAN, *et al.*, 2021).

3.4 Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations

O design da terapêutica antimicrobiana envolve a exploração de um vasto repertório químico para encontrar compostos com potência de amplo espectro e baixa toxicidade. Aqui, foi relatado um método computacional eficiente para a geração de antimicrobianos com atributos desejados. O método aproveita a orientação de classificadores treinados em um espaço latente informativo de moléculas modeladas usando um autoencoder generativo profundo e rastreiam as moléculas geradas usando classificadores de aprendizado profundo bem como características físico-químicas derivadas de simulações de dinâmica molecular de alto rendimento. Em 48 dias, foram identificados, sintetizados e testados experimentalmente 20 peptídeos antimicrobianos candidatos, dos quais dois apresentaram alta potência contra diversos patógenos Gram-positivos e Gram-negativos (incluindo Klebsiella pneumoniae multirresistente) e uma baixa propensão para induzir resistência a medicamentos em Escherichia coli. Ambos os peptídeos têm baixa toxicidade, conforme validado in vitro e em camundongos. Também foi mostrado, usando imagens confocais de células vivas, que o modo de ação bactericida dos peptídeos envolve a formação de membrana poros. A combinação de aprendizado profundo e dinâmica molecular pode acelerar a descoberta de potentes e seletivos antimicrobianos de amplo espectro (DAS, SERCU, *et al.*, 2021).

3.5 A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest X-ray images

Doenças pulmonares comuns são diagnosticadas pela primeira vez usando radiografias de tórax. Aqui, foi demonstrado que um pipeline de aprendizado profundo totalmente automatizado para a padronização das imagens de radiografia de tórax, para visualização de lesões e para diagnóstico de doenças pode identificar pneumonia viral causada pela doença de coronavírus 2019 (COVID-19) e avaliar sua gravidade, podendo também discriminar entre pneumonia viral causada pelo COVID-19 e outros tipos de pneumonia. O sistema de aprendizado profundo foi desenvolvido usando um conjunto de dados multicêntrico heterogêneo de 145.202 imagens e testado retrospectivamente e prospectivamente com milhares de imagens adicionais em todo o mundo. O sistema generalizou-se entre as configurações, discriminando entre

pneumonia viral, outros tipos de pneumonia e a ausência de doença com áreas sob a curva característica de operação do receptor (AUCs) de 0,94–0,98; entre COVID-19 grave e não grave com AUC de 0,87; e entre a pneumonia por COVID-19 e outras infecções virais ou pneumonia não viral com AUCs de 0,87–0,97. Em um conjunto independente de 440 radiografias de tórax, o sistema teve desempenho comparável aos radiologistas seniores e melhorou o desempenho dos radiologistas juniores. Sistemas automatizados de aprendizado profundo para avaliação de pneumonia podem facilitar a intervenção precoce e fornecer suporte para a tomada de decisões clínicas (WANG, LIU, *et al.*, 2021).

3.6 Shifting machine learning for healthcare from development to deployment and from models to data

Na última década, a aplicação de aprendizado de máquina na área da saúde ajudou a impulsionar a automação de tarefas médicas como bem como melhorias nas capacidades clínicas e acesso aos cuidados. Este progresso enfatizou que, desde o desenvolvimento do modelo até implantação do modelo, os dados desempenham funções centrais. Nesta revisão, foi fornecida uma visão centrada em dados das inovações e desafios que estão definindo o aprendizado de máquina para cuidados de saúde. Foram discutidos modelos generativos profundos e aprendizagem federada como estratégias para aumentar conjuntos de dados para melhorar o desempenho do modelo, bem como o uso dos modelos de transformadores mais recentes para lidar com conjuntos de dados maiores e aprimorando a modelagem do texto clínico. Também foram discutidos problemas com foco em dados na implantação de ML, enfatizando a necessidade para fornecer dados de forma eficiente a modelos de ML para previsões clínicas oportunas e para levar em consideração mudanças naturais de dados que podem se deteriorar desempenho do modelo (ZHANG, XING, *et al.*, 2022).

3.7 From predictive modelling to machine learning and reverse engineering of colloidal self-assembly

Uma diversidade esmagadora de blocos de construção coloidais com tamanhos distintos, materiais e potenciais de interação ajustáveis são agora disponíveis para automontagem coloidal. O espaço de aplicação para materiais compostos por esses blocos de construção é vasto. Fazer progresso no design racional de novos materiais

automontados, é desejável orientar os esforços de síntese experimental por modelagem computacional. Aqui, foram discutidos métodos de simulação de computador e estratégias usadas para o design de materiais macios criado através da automontagem ascendente de coloides e nanopartículas. Descrevemos técnicas de simulação para investigar o comportamento de automontagem de suspensões coloidais, incluindo métodos de previsão de estrutura cristalina, cálculos de diagrama de fase e técnicas de amostragem aprimoradas, bem como suas limitações. Também foi discutidos o recente aumento de interesse em aprendizado de máquina e métodos de engenharia reversa. Embora sua implementação no reino coloidal ainda esteja em sua infância, prevemos que essas ferramentas de ciência de dados oferecem novos paradigmas na compreensão, previsão e design (inverso) de novos materiais coloidais (DIJKSTRA e LUIJTEN, 2021).

3.8 Deep learning in histopathology: the path to the clinic.

As técnicas de aprendizado avançado têm grande potencial para melhorar o diagnóstico médico, oferecendo maneiras de melhorar a precisão, a reprodutibilidade e a velocidade, além de facilitar a carga de trabalho dos médicos. No campo da histopatologia, os algoritmos de aprendizado profundo foram desenvolvidos que executam de forma semelhante a patologistas treinados para tarefas como detecção e classificação de tumores. No entanto, apesar destes resultados promissores, muito poucos algoritmos alcançaram implementação clínica, desafiando o equilíbrio entre esperança e exageros para essas novas técnicas. Esta revisão fornece uma visão geral do estado atual do campo, bem como descreve os desafios que ainda precisam ser abordados antes que a inteligência artificial em histopatologia possa alcançar valor clínico (VAN DER LAAK, LITJENS e CIOMPI, 2021).

3.9 Deep learning for cellular image analysis

Avanços recentes em visão computacional e aprendizado de máquina sustentam uma coleção de algoritmos com uma capacidade impressionante de decifrar o conteúdo das imagens. Esses algoritmos de aprendizado profundo estão sendo aplicados a imagens biológicas e estão transformando a análise e interpretação de dados de imagem. Esses avanços estão posicionados para tornar as análises difíceis rotineiras e para permitir que os pesquisadores realizem novos experimentos antes impossíveis. Aqui, foi revisada a interseção entre aprendizagem e análise de

imagem celular e fornecida uma visão geral da mecânica matemática e da programação estruturas de aprendizado profundo que são pertinentes aos cientistas da vida. Pesquisou-se o progresso do campo em quatro aplicações principais: imagem classificação, segmentação de imagens, rastreamento de objetos e microscopia aumentada. Por fim, foi transmitida a experiência laboratorial com três aspectos principais da implementação do aprendizado profundo no laboratório: anotar dados de treinamento, selecionar e treinar um intervalo de arquiteturas de redes neurais e implantação de soluções. Também foram destacados conjuntos de dados e implementações existentes para cada aplicativo pesquisado (MOEN, BANNON, *et al.*, 2019).

3.10 Low-N protein engineering with data-efficient deep learning

A engenharia de proteínas tem um enorme potencial acadêmico e industrial. No entanto, é limitado pela falta de ensaios experimentais que são consistentes com o objetivo do projeto e com rendimento suficientemente alto para encontrar variantes raras e aprimoradas. Aqui, foi apresentado um paradigma guiado por aprendizado de máquina que pode usar apenas 24 sequências mutantes testadas funcionalmente para construir um virtual preciso cenário de fitness e tela de dez milhões de sequências por meio de evolução dirigida in silico. Como demonstrado em duas proteínas diferentes, GFP de Aequorea victoria (avGFP) e cepa de E. coli TEM-1 β-lactamase, os principais candidatos de uma única rodada são diversos e tão ativos quanto os mutantes projetados obtidos de esforços anteriores de alto rendimento. Ao destilar informações de proteínas naturais paisagens de sequência, nosso modelo aprende uma representação latente de 'não naturalidade', o que ajuda a guiar a pesquisa para longe de vizinhanças de sequência não funcionais. A supervisão subseqüente de baixo N identifica melhorias na atividade de interesse. Em suma, nossa abordagem permite o uso eficiente de ensaios de alta fidelidade com uso intensivo de recursos sem sacrificar o rendimento e ajuda a acelerar proteínas projetadas no fermentador, campo e clínica (BISWAS, KHIMULYA, *et al.*, 2021).

3.11 Precision Radiology: Predicting longevity using feature engineering and deep learning methods in a radiomics framework

As abordagens da medicina de precisão dependem da obtenção de conhecimento preciso do verdadeiro estado de saúde de um paciente individual, que

resulta de uma combinação de seus riscos genéticos e exposições ambientais. Esta abordagem é atualmente limitada pela falta de testes médicos não invasivos eficazes e eficientes para definem toda a gama de variação fenotípica associada à saúde individual. Tal conhecimento é crítico para melhorar a intervenção precoce, para melhores decisões de tratamento e para melhorar a constante agravamento da epidemia de doenças crônicas. Foram apresentados experimentos de prova de conceito para demonstrar como a imagem de tomografia computadorizada transversal adquirida rotineiramente pode ser usada para prever a longevidade do paciente como um substituto para estado geral de saúde e doença individual usando técnicas de análise de imagem de computador. Apesar do limitações de um conjunto de dados modesto e o uso de métodos de aprendizado de máquina disponíveis no mercado, nossos resultados são comparáveis aos métodos clínicos "manuais" anteriores para previsão da longevidade. Este trabalho demonstra que as técnicas de radiômica podem ser usadas para extrair biomarcadores relevantes para um dos mais amplamente utilizados resultados em pesquisas epidemiológicas e clínicas – mortalidade, e esse aprendizado profundo com convolução as redes neurais podem ser aplicadas de forma útil à pesquisa radiômica. Análise de imagem de computador aplicada a imagens médicas coletadas rotineiramente oferece um potencial substancial para aprimorar a medicina de precisão iniciativas (OAKDEN-RAYNER, CARNEIRO, *et al.*, 2017).

3.12 Weakly-supervised learning for lung carcinoma classification using deep learning

O câncer de pulmão é uma das principais causas de mortes relacionadas ao câncer em muitos países ao redor do mundo, e seu diagnóstico histopatológico é crucial para decidir sobre as melhores estratégias de tratamento. Recentemente, os modelos de aprendizado profundo de Inteligência Artificial têm se mostrado amplamente úteis em várias áreas médicas. campos, especialmente diagnósticos de imagem e patológicos; no entanto, modelos de IA para o diagnóstico patológico de lesões pulmonares que foram validadas em conjuntos de teste em larga escala ainda não foram vistas. Nós treinamos uma Rede Neural de Convolução baseada na arquitetura EfcientNet-B3, usando aprendizado de transferência e aprendizado fracamente supervisionado, para prever o carcinoma em imagens de slides inteiros (ISIs) usando um treinamento conjunto de dados de 3.554 ISIs. Foram obtidos

resultados altamente promissores para diferenciar carcinoma pulmonar e não neoplásica com altas AUCs em quatro conjuntos de teste independentes (AUCs de 0,975, 0,974, 0,988 e 0,981, respectivamente). Desenvolvimento e a validação de algoritmos como o apresentado são etapas iniciais importantes no desenvolvimento de suítes de software que poderiam ser adotados em práticas patológicas de rotina e potencialmente ajudar a reduzir a carga sobre patologistas (KANAVATI, TOYOKAWA, *et al.*, 2020).

3.13 Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning

A ausculta tem sido parte essencial do exame físico; isso é não invasivo, em tempo real, e muito informativo. A detecção de sons respiratórios anormais com um estetoscópio é importante para diagnóstico de doenças respiratórias e prestação de primeiros socorros. No entanto, a interpretação precisa dos sons respiratórios requer experiência considerável do clínico, então estagiários e residentes, às vezes, identificam erroneamente os sons respiratórios. Para superar tais limitações, tentamos desenvolver um sistema automatizado classificação dos sons respiratórios. Utilizamos a rede neural convolucional (CNN) de aprendizado profundo para categorizar 1.918 sons respiratórios (normais, crepitações, sibilos, roncos) registrados no contexto. Foi desenvolvido o modelo preditivo para classificação dos sons respiratórios combinando extrator de recursos de imagem de série, som respiratório e classificador CNN. Ele detectou sons anormais com acurácia de 86,5% e área sob a curva ROC (AUC) de 0,93 e classificou, ainda, sons pulmonares anormais em crepitações, sibilos ou roncos com uma precisão geral de 85,7% e uma média AUC de 0,92. Por outro lado, como resultado da classificação dos sons respiratórios por diferentes grupos apresentaram grau variável em termos de precisão; as acurácias gerais foram de 60,3% para estudantes de medicina,53,4% para estagiários, 68,8% para residentes e 80,1% para bolsistas. A classificação baseada em aprendizagem profunda seria capaz de complementar as imprecisões da ausculta dos médicos e pode ajudar no rápido diagnóstico e tratamento adequado das doenças respiratórias (KIM, HYON, *et al.*, 2021).

3.14 Integrating machine learning and multiscale modeling— perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences

Alimentadas por desenvolvimentos tecnológicos inovadores, as ciências biológicas, biomédicas e comportamentais estão agora coletando mais dados do que nunca. Há uma necessidade crítica de estratégias eficientes em termos de tempo e custo para analisar e interpretar esses dados para avançar saúde humana. A recente ascensão do aprendizado de máquina como uma técnica poderosa para integrar multimodalidade, dados de multifidelidade e revelar correlações entre fenômenos interligados apresenta uma oportunidade especial a esse respeito. No entanto, o aprendizado de máquina sozinho ignora as leis fundamentais da física e pode resultar em problemas mal colocados ou soluções não físicas. A modelagem multiescala é uma estratégia bem-sucedida para integrar dados multiescala e multifísica e descobrir mecanismos que explicam o surgimento da função. No entanto, a modelagem multiescala sozinha muitas vezes falha em combinar com eficiência grandes conjuntos de dados de diferentes fontes e diferentes níveis de resolução. Aqui, foi demonstrado que o aprendizado de máquina e a modelagem multiescala podem se complementar naturalmente para criar modelos preditivos robustos que integram a física subjacente para gerenciar problemas mal colocados e explorar espaços de design massivos. Analisou-se a literatura atual, destacaram-se aplicações e oportunidades, abordaram-se questões em aberto e discutiram-se possíveis desafios e limitações em quatro áreas tópicas abrangentes: equações diferenciais ordinárias, equações diferenciais parciais,abordagens e abordagens baseadas em teoria. Para atingir esses objetivos, foi aproveitada a experiência em matemática aplicada, ciência da computação, biologia computacional, biofísica, biomecânica, engenharia mecânica, experimentação e medicina. A perspectiva multidisciplinar sugere que a integração de aprendizado de máquina e modelagem multiescala pode fornecer novos insights sobre mecanismos de doenças, ajudando a identificar novos alvos e estratégias de tratamento e informando a tomada de decisões em benefício da saúde humana (ALBER, BUGANZA TEPOLE, *et al.*, 2019).

3.15 Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis

O aprendizado profundo (DL) tem o potencial de transformar os diagnósticos médicos. No entanto, a precisão diagnóstica do DL é incerta. Nosso foco foi avaliar a precisão diagnóstica dos algoritmos DL para identificar patologias em imagens médicas. As buscas foram realizadas em Medline e EMBASE até janeiro de 2020. Identificamos 11.921 estudos, dos quais 503 foram incluídos na revisão sistemática. Oitenta e dois estudos em oftalmologia, 82 em doenças da mama e 115 em doenças respiratórias foram incluídos para meta-análise. Duzentos vinte e quatro estudos em outras especialidades foram incluídos para revisão qualitativa. Estudos revisados por pares que relataram sobre o diagnóstico precisão de algoritmos DL para identificar patologia usando imagens médicas foram incluídos. Os resultados primários foram medidas de precisão diagnóstica, desenho do estudo e padrões de relatórios na literatura. As estimativas foram agrupadas usando meta-análise de efeitos aleatórios. Em oftalmologia, a AUC variou entre 0,933 e 1 para diagnosticar retinopatia diabética, macular relacionada à idade, degeneração e glaucoma em fotografias de fundo de retina e tomografia de coerência óptica. Em imagens respiratórias, AUC's variou entre 0,864 e 0,937 para o diagnóstico de nódulos pulmonares ou câncer de pulmão na radiografia de tórax ou tomografia computadorizada. Para imagens de mama, AUC's variou entre 0,868 e 0,909 para o diagnóstico de câncer de mama em mamografia, ultrassonografia, ressonância magnética e tomossíntese digital da mama. A heterogeneidade foi alta entre os estudos e foi observada uma extensa variação na metodologia, terminologia e medidas de resultados. Isso pode levar a uma superestimação da precisão diagnóstica dos algoritmos DL em imagens médicas. Há uma necessidade imediata de o desenvolvimento de diretrizes EQUATOR específicas para inteligência artificial, particularmente STARD, a fim de fornecer orientações sobre as principais questões neste campo (AGGARWAL, SOUNDERAJAH, *et al.*, 2021).

3.16 Radiomics-guided deep neural networks stratify lung adenocarcinoma prognosis from CT scans

Deep Learning (DL) é uma tecnologia inovadora para imagens médicas com alto tamanho de amostra requisitos e questões de interpretabilidade. Usando um modelo DL pré-treinado por meio de uma abordagem guiada por radiômica, propomos

uma metodologia para estratificar o prognóstico de adenocarcinomas de pulmão com base na tomografia computadorizada pré-tratamento. Esta abordagem permitiu aplicar DL com menores requisitos de tamanho de amostra e interpretabilidade aprimorada. Radiômica de linha de base e modelos DL para o prognóstico de adenocarcinomas pulmonares foram desenvolvidos e testados usando local (n = 617) coorte. Os modelos DL foram ainda testados em uma coorte de validação externa (n = 70). O conjunto foi dividido em conjuntos de treinamento e de teste. Uma pontuação de risco radiômica (Radiomic Risk Score - RRS) foi desenvolvida usando Cox-LASSO. Três redes DL pré-treinadas derivadas de imagens naturais foram usadas para extrair os recursos DL. Os recursos foram guiados ainda mais usando radiômica, mantendo aqueles As feições DL cujas correlações com as feições radiômicas foram altas e os valores-p corrigidos por Bonferroni foram baixos. Os recursos DL retidos foram sujeitos a um Cox-LASSO quando construção de escores de risco de DL. Os grupos de risco estratificados pelo RRS apresentaram diferença significativa em conjuntos de treinamento, teste e validação. Os recursos DL foram interpretados usando recursos radiômicos existentes, e os recursos de textura explicaram o DL apresenta bem (CHO, LEE, *et al.*, 2021).

4 METODOLOGIA

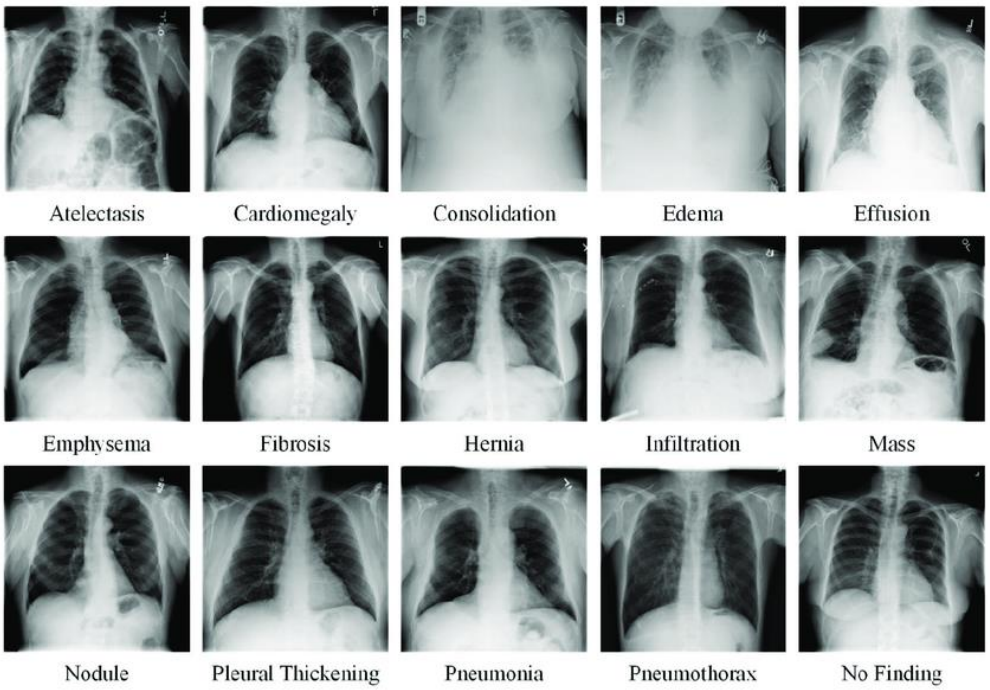
A metodologia desta pesquisa pode ser dividida em três aspectos: coleta de dados, seleção da arquitetura da rede neural e da métrica mais adequada para avaliar o treinamento do modelo e a execução do treinamento de acordo com a metodologia e critérios selecionados durante a pesquisa.

4.1 Banco de dados

Primeiramente, houve uma preocupação com a quantidade de dados disponíveis para treinamento da rede neural. Como mencionado nos capítulos anteriores desta dissertação, modelos mais complexos (profundos) de aprendizado de máquina precisam, necessariamente, de mais dados do que modelos menores (rasos). Isso se dá pela própria característica construtiva dos modelos de *deep learning*: quanto mais nós (neurônios) uma rede neural possui, maior é a chance de ocorrer *overfitting*. Em outras palavras, se uma rede neural muito complexa for treinada com poucos dados, na prática, o que ela faz é “decorar” os dados de treinamento, ao invés de extrair características capazes de generalizar o processo de classificação para dados desconhecidos.

Face à necessidade de dados robustos em grande quantidade, foi escolhido o banco de dados público ChestX-ray14 (WANG, PENG, *et al.*, 2017), que é um conjunto de imagens médicas que contém 112120 imagens de raio-x frontais de 30805 pacientes, coletadas entre 1992 e 2015. Todas as imagens são rotuladas com pelo menos uma de quinze classes, sendo uma classe para exames nos quais nenhuma doença foi identificada (60361 imagens) e as demais quatorze classes para as seguintes doenças: atelectasia (11559 imagens), cardiomegalia (2776 imagens), efusão (13317 imagens), infiltração (19894 imagens), massa (5782 imagens), nódulo (3661 imagens), pneumonia (1431 imagens), pneumotórax (5302 imagens), consolidação (4667 imagens), edema (2303 imagens), enfisema (2516 imagens), fibrose (1686 imagens), espessamento pleural (3385 imagens) e hérnia (227 imagens). Note-se que a soma das quantidades de imagens para cada classe resulta num número maior que 112120, justamente porque as classes de doenças não são exclusivas. Com exceção das 60361 imagens que não apresentam doenças, as demais 51759 não estão necessariamente limitadas a uma única classe. A Figura 3 traz exemplos das quinze classes presentes no banco de dados ChestX-ray14.

Figura 3: Exemplos do banco de dados ChestX-ray14.



Fonte: (WANG, PENG, *et al.*, 2017)

Todas as imagens, além de estarem rotuladas, também estão padronizadas no que diz respeito às dimensões: 1024x1024x1 pixels, sendo que o primeiro número indica a largura, o segundo se refere à altura e o terceiro é a quantidade de canais de cores (1 para monocromático ou 3 para RGB).

4.2 Arquitetura e métricas

Conforme observado na revisão bibliográfica desta pesquisa, o estado da arte para classificação de imagens com *deep learning* é, atualmente, a rede neural convolucional (CNN) e, por esse motivo, esse tipo de arquitetura foi escolhido para processar os dados mencionados na seção anterior. Contudo, as CNNs não são redes neurais com estrutura fixa. De fato, dizer que uma rede neural é uma CNN apenas implica na existência de camadas convolucionais dentro da arquitetura, sem especificar quantas são as camadas convolucionais e quais são as camadas intermediárias, informações essas que são determinadas segundo o critério do cientista de dados.

Na seção anterior, foi discutida a relação entre a quantidade de dados e o *overfitting* de uma rede neural. No entanto, a arquitetura e a metodologia de treinamento de uma rede neural também são fatores determinantes do quão

suscetível o modelo é ao *overfitting*. Para minimizar esse risco, há técnicas já consagradas, como a divisão do conjunto de dados em dois subconjuntos (treinamento e validação), a fim de que a cada ciclo de treinamento, denominado época (*epoch*), a rede neural possa otimizar os seus parâmetros com base no subconjunto de treinamento e verificar se os ajustes feitos durante um dado ciclo melhoraram ou pioraram a capacidade de generalização com base no subconjunto de validação. Por fim, há um terceiro conjunto, denominado conjunto de teste, que é utilizado após o término de todas as épocas de treinamento para determinar as métricas finais da rede neural.

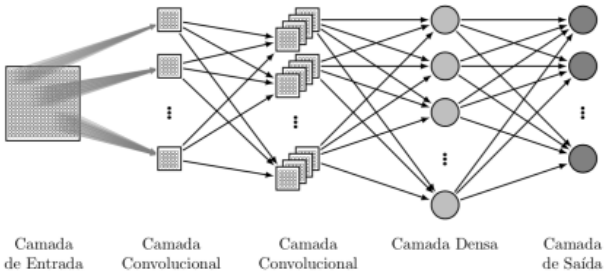
Outra técnica consiste na inserção de uma ou mais camadas de *dropout* na arquitetura da rede neural. Esse tipo de camada não possui parâmetros treináveis e sua única finalidade é bloquear, com base num valor percentual, conexões aleatórias entre os neurônios de duas camadas. Quando uma rede neural apresenta *overfitting*, em geral, pode-se dizer que os neurônios passaram a aprender características raras, ou seja, elementos que caracterizam exclusivamente os dados de treinamento, mas que não se aplicam aos demais dados. Como as redes neurais tendem a ter um comportamento previsível, pode-se dizer que tais características raras são detectadas sempre pelos mesmos neurônios. Ao aplicar uma camada de *dropout*, espera-se que esse processo seja inibido, pois, quando a conexão dos neurônios especializados nessas características for bloqueada, a rede neural deverá buscar outros caminhos para chegar à classificação correta.

Ambas as técnicas supracitadas foram consideradas na construção da rede neural utilizada nesta pesquisa.

Quanto à arquitetura propriamente dita, deve-se considerar o tipo de dado que a CNN receberá na camada de entrada e o tipo de informação que ela deverá entregar na saída. Sabe-se que os dados de entrada serão imagens com dimensões 1024x1024x1, mas, com base em trabalhos anteriores, observou-se que muitas análises semelhantes foram bem-sucedidas com imagens redimensionadas para 256x256x1, o que também é relevante do ponto de vista de eficiência de treinamento, dado que quanto mais *pixels* a imagem de entrada tiver, maiores serão a largura da camada de entrada, a quantidade de parâmetros treináveis, a quantidade de memória necessária no *hardware* e o tempo de treinamento.

Além desse aspecto, também é importante ter em mente o tipo de tarefa a ser executada. Neste caso, foi feita uma classificação das imagens, o que implica no acréscimo de camadas densas de neurônios, ou camadas totalmente conectadas, na saída da última camada convolucional. Nesse tipo de camada, todos os neurônios de uma camada são conectados a todos os neurônios da camada seguinte, de forma que as características extraídas pelas camadas convolucionais sejam combinadas a fim de classificar corretamente as imagens. A Figura 4 representa uma arquitetura genérica de uma rede neural convolucional classificadora.

Figura 4: Rede neural convolucional classificadora.



Fonte: (BAIA e CASTRO, 2018)

Uma vez definido o tipo da camada de saída, deve-se determinar suas dimensões e função de ativação. A quantidade de nós de saída depende da quantidade de classes que devem ser preditas: em uma predição puramente binária, ou seja, com valor 0 ou 1, a última camada da rede neural contém um único neurônio, porém, quando se faz uma classificação com mais classes, a camada de saída tem tamanho similar ao número de classes. Quanto à função de ativação, deve-se considerar se as classes são exclusivas ou não: quando se tem um conjunto de dados com classes exclusivas (*multi-class classification*), é comum utilizar a função de ativação *softmax*, mas, quando as classes não são exclusivas (*multi-label classification*), costuma-se utilizar a função de ativação *sigmoid*, que também é utilizada em casos de classificação binária (*binary classification*). No caso do conjunto de dados escolhidos para esta pesquisa, há duas possibilidades: fazer uma classificação binária com a classe exclusiva (imagem com ou sem sinais de doenças) ou uma classificação *multi-label* com as classes não exclusivas. Por questões de desbalanceamento entre as classes do conjunto de dados (há classes com quantidade de imagens muito pequena comparado com a totalidade do conjunto de dados), esta pesquisa focou na classificação binária, uma vez que se pode usar somente a classe

exclusiva para treinar um modelo com uma única saída binária, onde 0 significa que a imagem tem sinais de doenças e 1 significa que não foram detectadas doenças.

Quanto à avaliação de um modelo de rede neural, há várias métricas disponíveis nos *frameworks* de treinamento e, por esta pesquisa lidar com classificações binárias, serão apresentadas métricas que se baseiam na comparação entre os rótulos reais e as predições feitas pelo modelo, ou seja, há dois possíveis valores para os rótulos e dois possíveis valores de predição (0 = doente e 1 = saudável), possibilitando a construção de uma matriz de dimensões 2x2, denominada matriz de confusão, que correlaciona esses valores (Figura 5).

Figura 5: Matriz de confusão.

		Valor Predito	
		Sim	**Não**
Real	**Sim**	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Fonte: (NOGARE, 2020)

Conforme visto na Figura 5, a diagonal principal da matriz de confusão representa os acertos: se o valor real for 1 e o valor predito for 1, diz-se que o resultado é um verdadeiro positivo (True Positive – TP); se o valor real for 0 e o valor predito for 0, diz-se que o resultado é um verdadeiro negativo (True Negative – TN). Por outro lado, a diagonal secundária representa as predições incorretas: se o valor real for 1 e o valor predito for 0, denomina-se o resultado como falso negativo (False Negative – FN); se o valor real for 0 e o valor predito for 1, denomina-se o resultado como falso positivo (False Positive – FP).

Numa classificação binária, todas as métricas de avaliação se baseiam nesses quatro valores. Embora existam outras, a Tabela 2 apresenta cinco métricas com as respectivas fórmulas em função dos elementos da matriz de confusão.

Tabela 2: Métricas de avaliação.

Métrica	**Cálculo**
Acurácia	$\frac{TP+TN}{TP+TN+FP+FN}$
Precisão	$\frac{TP}{TP+FP}$
Recall/Sensibilidade	$\frac{TP}{TP+FN}$
Valor preditivo negativo	$\frac{TN}{TN+FN}$
Especificidade	$\frac{TN}{TN+FP}$

Fonte: Vaz Gabriel (2023)

A acurácia é a métrica mais conhecida e representa, dentre a totalidade das predições, quais foram feitas corretamente. Por essa razão, divide-se os elementos da diagonal principal da matriz de confusão pela soma de todos os elementos. Porém, embora seja muito utilizada, não se recomenda o seu uso em banco de dados desbalanceados. Se, por exemplo, um conjunto de dados possui 90% de dados pertencentes à classe positiva, basta o modelo predizer todos os dados como positivos e ele obterá uma acurácia de 90%, que, teoricamente, é um valor perfeitamente aceitável em modelos de redes neurais.

A precisão e a sensibilidade de uma rede neural dizem respeito à capacidade do modelo em predizer a classe positiva. A primeira relaciona todos os dados que foram correta ou incorretamente preditos como positivos (TP+FP) com aqueles que foram classificados corretamente (TP), ou seja, a precisão é o percentual de dados classificados como positivos que, de fato, o são. A segunda relaciona todos os dados cujos valores reais são positivos, independente da predição (TP+FN), com os valores classificados corretamente (TP), ou seja, a sensibilidade é o percentual de dados realmente positivos que foram classificados como tais.

O valor preditivo negativo e a especificidade são o oposto da precisão e da sensibilidade e medem a capacidade do modelo em predizer a classe negativa. A primeira relaciona todos os dados que foram correta ou incorretamente preditos como negativos (TN+FN) com aqueles que foram classificados corretamente (TN), ou seja, o valor preditivo negativo é o percentual de dados classificados como negativos que,

de fato, o são. A segunda relaciona todos os dados cujos valores reais são negativos, independente da predição (TN+FP), com os valores classificados corretamente (TN), ou seja, a especificidade é o percentual de dados realmente negativos que foram classificados como tais.

A métrica de avaliação de uma rede neural varia muito em função do tipo dos dados e do que o cientista de dados deseja extrair do modelo. No caso desta pesquisa, se está observando somente uma classe, que é o rótulo que indica se existe algum indício de doenças nas imagens, sendo que a classe positiva (valor 1) indica um paciente saudável e a classe negativa (valor 0) indica a detecção de doenças.

Para escolher a melhor métrica de avaliação para esta pesquisa, a primeira consideração foi o desbalanceamento do conjunto de dados (há mais imagens de pacientes saudáveis do que imagens de pacientes enfermos), mesmo após a separação do banco de dados em somente duas classes, ao invés das quinze originais. Essa característica inviabiliza o uso da acurácia, pois, como mencionado anteriormente, ela pode ocultar classificações incorretas na classe minoritária.

A segunda consideração feita para a escolha da métrica de avaliação foi o fato de que toda e qualquer rede neural apresenta erros, ou seja, não existem redes neurais capazes de acertar 100% das classificações. Dessa forma, deve-se entender os dados e determinar qual classe (positiva ou negativa) traz maior prejuízo se for predita incorretamente. No caso desta pesquisa, que está envolvida com a área médica, determinou-se que as predições corretas da classe negativa são mais importantes, pois é mais aceitável classificar incorretamente um paciente saudável do que um doente. Por “aceitável”, entenda-se que os modelos de aprendizado de máquina, principalmente em situações que envolvem risco de vida, devem sempre complementar a avaliação de um especialista e nunca devem ser utilizados de forma independente: uma classificação de um paciente saudável como enfermo pode ser facilmente revista por um profissional da área médica.

Dessa forma, após considerar os itens mencionados anteriormente, determinou-se que as melhores métricas para avaliação da rede neural utilizada nesta pesquisa são aquelas que avaliam a classe negativa, ou seja, valor preditivo negativo e especificidade.

4.3 Treinamento

Para o treinamento da rede neural, foi utilizado um computador com processador Intel® Core™ i7-11800H @ 2.30 GHz de 11ª geração, memória RAM de 16 GB, SSD de 512 GB e GPU Nvidia® GeForce RTX™ 3060 com 6 GB de memória dedicada.

O ambiente de programação foi a edição individual do Anaconda®, que é uma plataforma *open source* dedicada à ciência dos dados. A linguagem de programação utilizada na construção da rede neural, separação do conjunto de dados nos subconjuntos de treinamento, validação e teste, treinamento da rede neural e teste da mesma foi o *Python* versão 3.9.7 dentro do ambiente de desenvolvimento interativo *JupyterLab* versão 3.5.2. Para executar essas etapas, algumas bibliotecas disponibilizadas pelo repositório do Anaconda® foram usadas: *pandas* versão 1.3.5, para extrair os rótulos das imagens do arquivo CSV disponibilizado junto com o conjunto de dados e organizar a separação dos subconjuntos de dados; *tensorflow* versão 2.5.0, que é uma biblioteca dedicada ao aprendizado de máquina e, atualmente, tem integrado o *keras*, que é uma biblioteca com módulos específicos para *deep learning* e redes neurais. Todas as camadas, modelos, otimizadores, aplicações, funções de retorno e métricas utilizados na construção e compilação da rede neural estão presentes nesta última biblioteca.

Conforme foi descrito anteriormente, uma rede neural convolucional classificadoras é formada por um conjunto de camadas convolucionais, responsáveis pela extração de características dos dados de entrada, e um conjunto de camadas densas, responsáveis pela classificação propriamente dita. Dado que o treinamento das camadas convolucionais exige muito poder computacional e muito tempo, mesmo num computador com recursos dedicados, como o utilizado, preferiu-se lançar mão de um recurso chamado *transfer learning*, que consiste no aproveitamento de uma rede neural, ou pelo menos parte de uma rede neural, pré-treinada em outro conjunto de dados.

Dado o histórico de aplicações presentes na literatura médica, optou-se por utilizar parte da rede neural VGG-16, que é uma rede neural com cinco blocos convolucionais, totalizando dezoito camadas (treze convolucionais e cinco de *pooling*),

e um bloco de três camadas densas (SIMONYAN e ZISSERMAN, 2015), conforme mostra a Figura 6.

Figura 6: Arquitetura VGG-16.

Fonte: (ROHINI, 2021)

A rede VGG-16 foi treinada com o conjunto de imagens *ImageNet* que, à época, continha três grupos de imagens (treinamento, com 1,3 milhões de imagens, validação, com 50 mil imagens, e teste, com 100 mil imagens) pertencentes a 1000 classes e obteve o primeiro lugar no desafio *ImageNet* de 2014, obtendo resultados significativos com outros conjuntos de dados, demonstrando a capacidade de generalização dessa rede (SIMONYAN e ZISSERMAN, 2015).

Nesta pesquisa, foi utilizada a base convolucional da rede VGG-16, porém as camadas classificadoras, que fazem a classificação propriamente dita foram substituídas por novas camadas que foram treinadas do zero. É importante ressaltar que foram feitas duas etapas de treinamento: uma etapa com a base convolucional totalmente congelada, de modo que os pesos pré-treinados não fossem alterados, e outra etapa com treinamento apenas de algumas camadas do último bloco, a fim de fazer um ajuste fino do aprendizado. A primeira etapa visa treinar exclusivamente as camadas densas com as características extraídas pela base convolucional da rede VGG-16, uma vez que não há necessidade de retreinar as camadas convolucionais para extrair padrões comuns, como linhas e formas geométricas. Contudo, as camadas mais profundas da base convolucional são responsáveis pela extração de características abstratas, que tendem a ser mais exclusivas ao conjunto de imagens com as quais elas foram treinadas. Por essa razão, após a primeira etapa, quando as camadas classificadoras já estão otimizadas, habilita-se o treinamento do último bloco convolucional para que ele se ajuste para extrair as características abstratas do conjunto de dados utilizados na pesquisa.

Para o treinamento da rede neural, foram criadas duas funções de retorno que têm a finalidade de otimizar o treinamento em tempo real. Na prática, elas avaliam a evolução das métricas de treinamento a cada época e tomam decisões no que diz respeito aos hiper parâmetros da rede, a fim de evitar o *overfitting*. A primeira função monitora a perda da rede no conjunto de validação, a fim de identificar o momento em que a rede deixa de aprender (o chamado platô de aprendizado), e, quando esse momento é identificado, ela reduz a taxa de aprendizado do treinamento, que é o quanto os pesos são modificados de acordo com o gradiente de perda, na tentativa de fazer a rede buscar o valor mínimo de perda. A segunda função monitora os verdadeiros negativos do conjunto de validação com a finalidade de identificar a partir de qual época de treinamento a rede deixa de melhorar de forma consistente a principal métrica de avaliação. Ao contrário da primeira função, esta interrompe o treinamento, pois, quando a métrica de validação deixa de evoluir, é um possível sinal de que a rede está prestes a entrar numa condição de *overfitting*, de forma que é melhor interromper o treinamento e salvar a rede com os pesos obtidos até então, do que permitir que a continuação do treinamento degrade os pesos.

Para a execução do treinamento, o banco de dados foi dividido conforme sugerido pelo fornecedor das imagens, totalizando 54,3% para conjunto de treinamento, 22,8% para conjunto de validação e 22,8% para conjunto de teste, sendo que todas as imagens foram normalizadas para que os *pixels* tivessem valores entre 0 e 1. O treinamento foi programado para ser realizado em vinte épocas, embora houvesse a possibilidade de treinar em menos, caso as funções de retorno atuassem nesse sentido.

5 ANÁLISE DOS RESULTADOS

Os resultados que serão apresentados neste capítulo são as métricas de treinamento e o resultado final do modelo no conjunto de teste. Para fins de comparação, foram treinadas duas redes neurais de acordo com os procedimentos descritos no capítulo anterior, porém, com diferentes configurações na segunda etapa de treinamento. Na primeira rede neural, o treinamento durante a segunda etapa foi feito com o quinto bloco convolucional (camadas Conv 5-1, Conv 5-2 e Conv 5-3) completamente descongelado e sem funções de retorno. Na segunda rede neural, o treinamento da segunda etapa foi feito somente com a última camada convolucional (Conv 5-3) descongelada e com ambas as funções de retorno mencionadas anteriormente.

A arquitetura das redes neurais durante a primeira etapa dos testes, com todas as camadas da rede VGG-16 congeladas, está representada na Tabela 3. Note-se que, embora a arquitetura na primeira etapa seja a mesma em ambas as redes treinadas, dois treinamentos completos foram executados, sem que os pesos otimizados na primeira etapa dos treinamentos fossem compartilhados entre as redes. Isso se deve ao fato de que houve estratégias diferentes nos treinamentos no que diz respeito ao uso das funções de retorno.

Tabela 3: Arquitetura da rede neural na primeira etapa de treinamento.

Camada	**Quantidade de parâmetros**	**Parâmetros treináveis**
VGG-16	14.714.688	0
Flatten	0	0
Dense	16.777.728	16.777.728
Dropout	0	0
Dense	513	513

Fonte: Vaz Gabriel (2023)

Visto que, nesta etapa, a rede VGG-16 está congelada, os únicos parâmetros treináveis são os 16.778.241 parâmetros referentes às camadas densas.

5.1 Rede neural com três camadas convolucionais treináveis

A Tabela 4 apresenta a arquitetura da rede neural adotada para o primeiro treinamento, com todas as camadas convolucionais do último bloco da rede VGG-16 habilitadas para treinamento, totalizando 23.857.665 parâmetros treináveis.

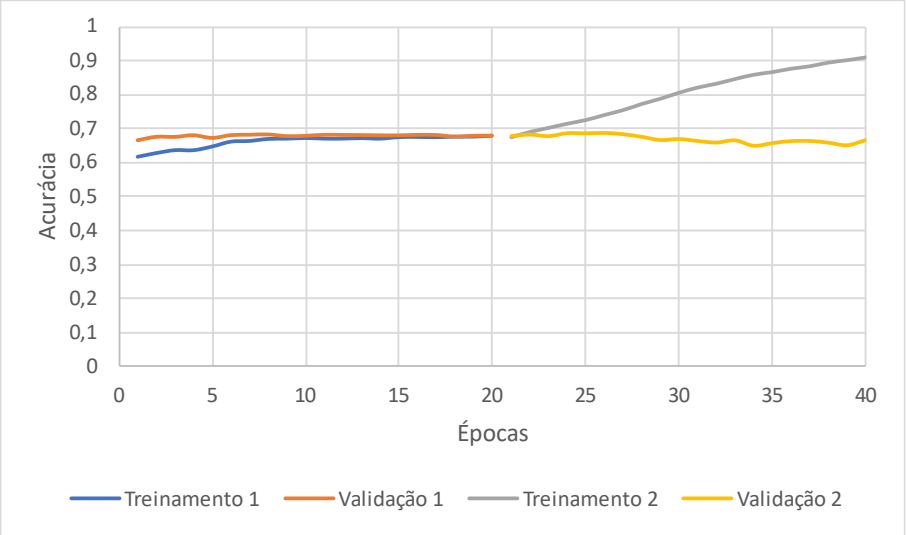
Tabela 4: Arquitetura da rede neural com três camadas convolucionais treináveis.

Camada	Quantidade de parâmetros	Parâmetros treináveis
VGG-16	14.714.688	7.079.424
Flatten	0	0
Dense	16.777.728	16.777.728
Dropout	0	0
Dense	513	513

Fonte: Vaz Gabriel (2023)

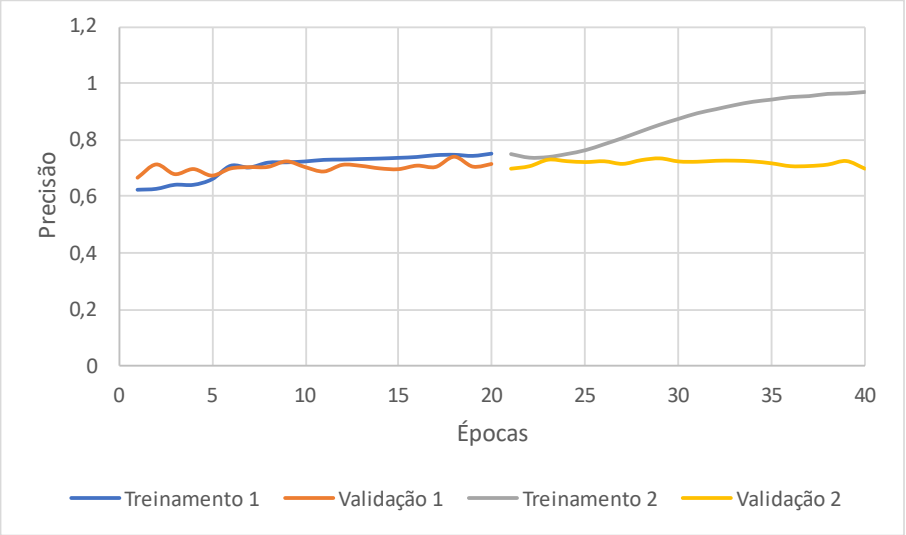
Neste treinamento, todas as 40 épocas de treinamento foram executadas (20 para cada etapa) e, embora o foco desta pesquisa seja a especificidade do modelo, todas as cinco métricas discutidas anteriormente serão apresentadas (Figuras 7 a 11), a fim de permitir uma compreensão abrangente do desempenho da rede neural no decorrer do treinamento.

Figura 7: Acurácia da CNN com três camadas convolucionais treináveis.



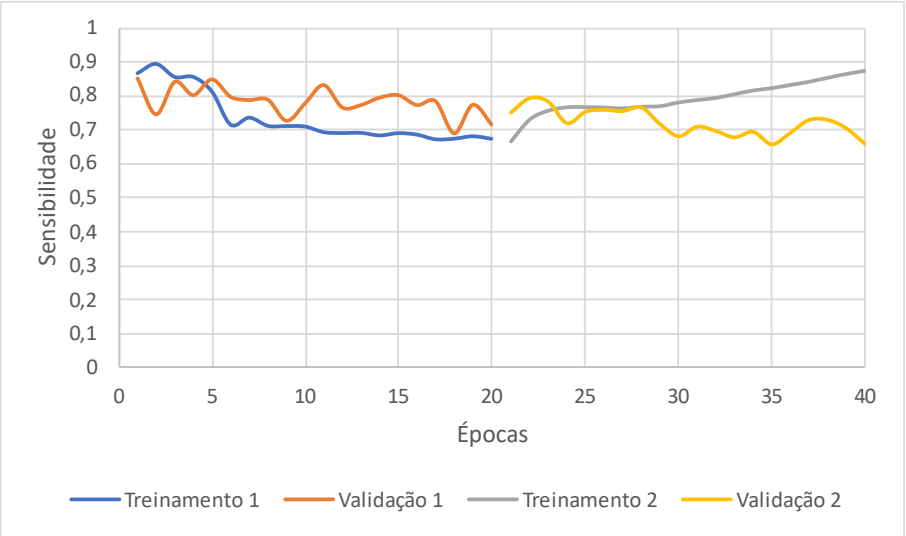
Fonte: Vaz Gabriel (2023)

Figura 8: Precisão da CNN com três camadas convolucionais treináveis.



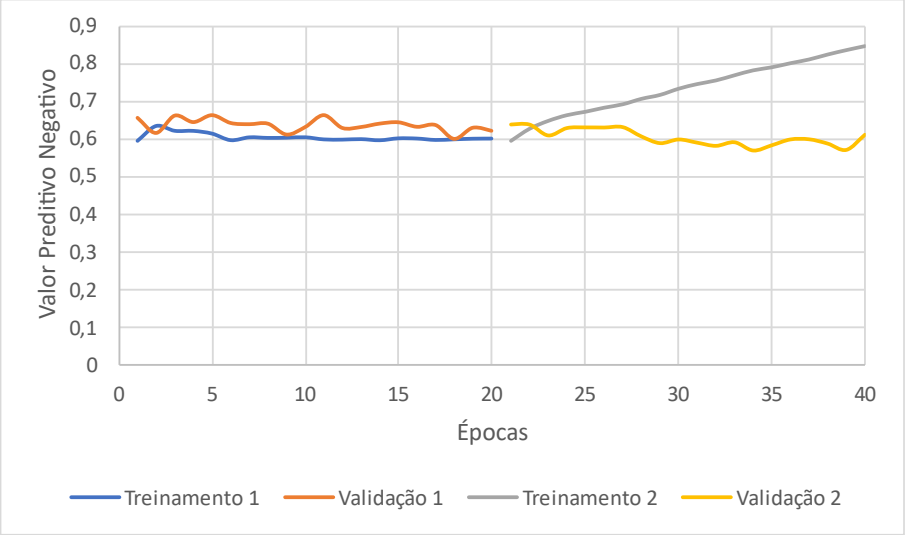
Fonte: Vaz Gabriel (2023)

Figura 9: Sensibilidade da CNN com três camadas convolucionais treináveis.



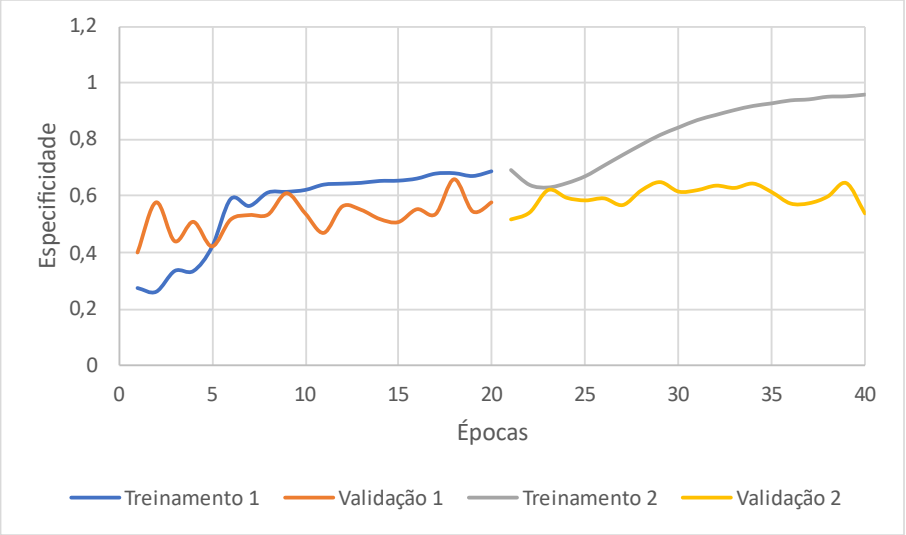
Fonte: Vaz Gabriel (2023)

Figura 10: Valor preditivo negativo da CNN com três camadas convolucionais treináveis.



Fonte: Vaz Gabriel (2023)

Figura 11: Especificidade da CNN com três camadas convolucionais treináveis.



Fonte: Vaz Gabriel (2023)

Todos os gráficos apresentam a evolução das métricas de treinamento e validação nas duas etapas de treinamento e alguns pontos importantes devem ser observados.

A primeira observação é que a acurácia, de fato, não é uma métrica adequada para este tipo de aplicação, uma vez que, tanto no grupo de treinamento quanto no de validação, ela permaneceu praticamente constante durante toda a primeira etapa

da otimização da rede. Por outro lado, as demais variáveis, principalmente a sensibilidade e a especificidade, tiveram grandes oscilações.

A segunda observação diz respeito justamente à correlação entre a sensibilidade e a especificidade no conjunto de validação. Ao passo que a sensibilidade começou o treinamento com um valor próximo a 0,9 e, posteriormente, diminuiu para valores entre 0,7 e 0,8, a especificidade começou com um valor próximo a 0,4 e aumentou para 0,6, onde se estabilizou ao final da primeira etapa de treinamento. Esse efeito mostra que, no início do treinamento, o modelo simplesmente "chutava" as classificações com base na classe majoritária (positiva). Isso fica evidente quando se analisa a precisão, que começou próxima de 0,6 e, conforme a sensibilidade diminuiu, ela se estabilizou entre 0,7 e 0,8. Isso indica que, no início, a maior parte da classe positiva era classificada corretamente (sensibilidade alta), porém, aproximadamente 40% das classificações positivas eram falsos positivos. O efeito inverso ocorre simultaneamente com a especificidade: conforme o modelo era treinado, ele se tornou menos sensível à classe positiva e passou a classificar corretamente uma maior quantidade de dados da classe negativa. Por outro lado, o valor preditivo negativo permaneceu praticamente constante durante todo o treinamento, indicando que a quantidade de falsos negativos aumentou, afinal, se o conjunto total de classificações negativas aumentou (especificidade mais alta) e a proporção de falsos negativos permaneceu constante, logicamente, a quantidade de falsos positivos também aumentou.

A terceira observação diz respeito à segunda etapa do treinamento (épocas 21 a 40). Em todos os gráficos, se observa uma tendência clara de *overfitting*, uma vez que todas as métricas do conjunto de treinamento se distanciam das métricas do conjunto de validação a partir da 23ª época, atingindo valores próximos de 0,9 e 1. Isso indica que o modelo passou a memorizar o conjunto de treinamento, perdendo a capacidade de generalização e estagnando a evolução das métricas para conjuntos desconhecidos. Isso fica claro quando se compara as métricas obtidas para o conjunto de teste, que estão apresentadas na Tabela 5.

Tabela 5: Métricas do conjunto de teste para o primeiro modelo treinado.

Etapa	Acurácia	Precisão	Sensibilidade	Valor preditivo negativo	Especificidade
1	0,67	0,59	0,45	0,70	0,81
2	0,64	0,54	0,52	0,71	0,72

Fonte: Vaz Gabriel (2023)

A evolução dessas métricas, entre a primeira e a segunda etapa de treinamento, indica claramente a degradação da generalização do modelo, uma vez que ele se tornou, novamente, mais sensível à classe positiva (majoritária), com um aumento da sensibilidade de 0,45 para 0,52, em detrimento da classificação correta da classe negativa (minoritária), com diminuição da especificidade de 0,81 para 0,72.

5.2 Rede neural com uma camada convolucional treinável

A Tabela 6 apresenta a arquitetura da rede neural adotada para o segundo treinamento, com somente uma camada convolucional do último bloco da rede VGG-16 habilitada para treinamento, totalizando 19.138.049 parâmetros treináveis.

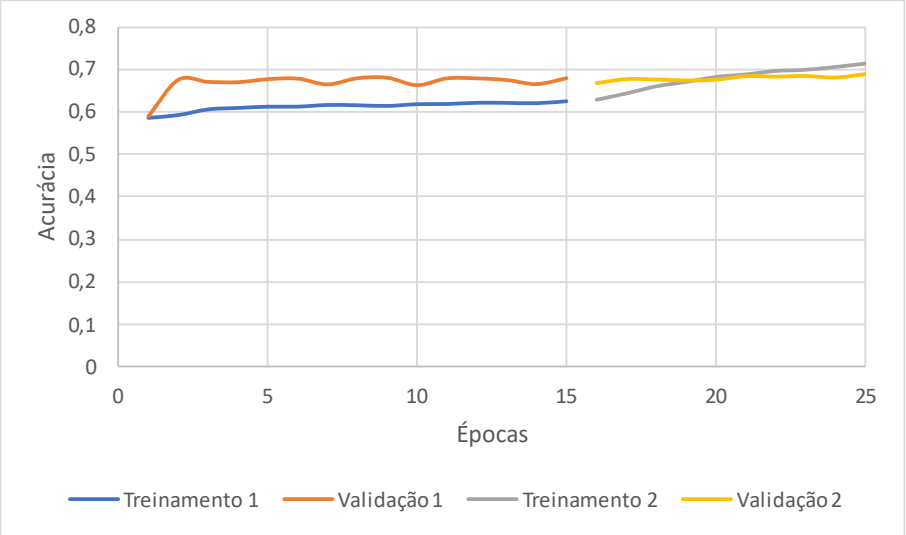
Tabela 6: Arquitetura da rede neural com uma camada convolucional treinável.

Camada	Quantidade de parâmetros	Parâmetros treináveis
VGG-16	14.714.688	2.359.808
Flatten	0	0
Dense	16.777.728	16.777.728
Dropout	0	0
Dense	513	513

Fonte: Vaz Gabriel (2023)

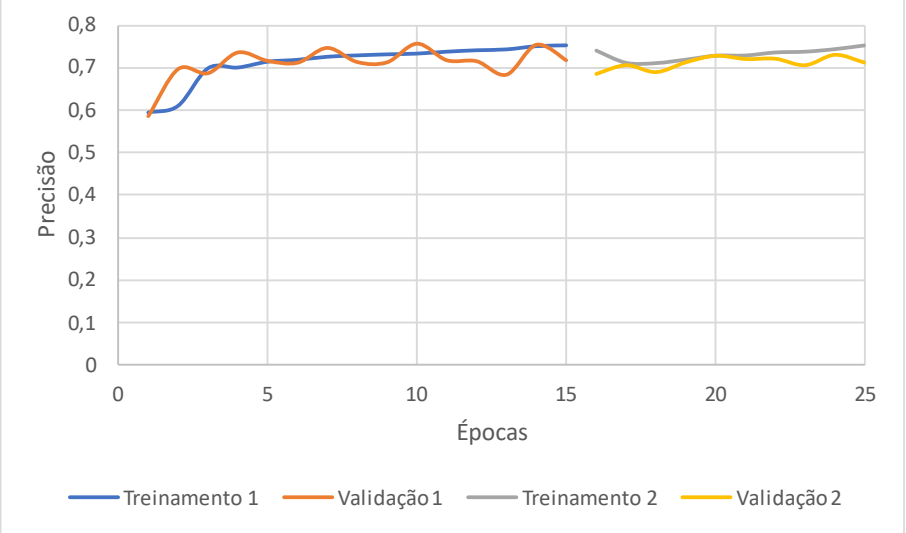
Neste treinamento, foram aplicadas as funções de retorno com capacidade de ajustar a taxa de aprendizado do modelo em tempo real e interromper o treinamento se for entendido que já foram obtidos os melhores resultados. Dessa forma, houve somente 15 épocas de treinamento na primeira etapa e 10 épocas de treinamento na segunda, totalizando 25 épocas, ao invés das 40 pré-determinadas. Novamente, todas as cinco métricas discutidas anteriormente serão apresentadas (Figuras 12 a 16), a fim de permitir uma compreensão abrangente do desempenho da rede neural no decorrer do treinamento.

Figura 12: Acurácia da CNN com uma camada convolucional treinável.



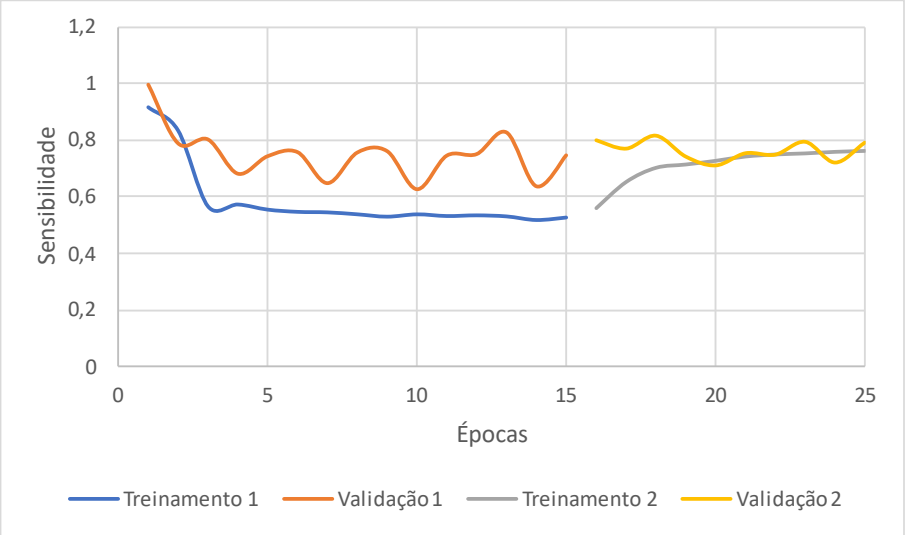
Fonte: Vaz Gabriel (2023)

Figura 13: Precisão da CNN com uma camada convolucional treinável.



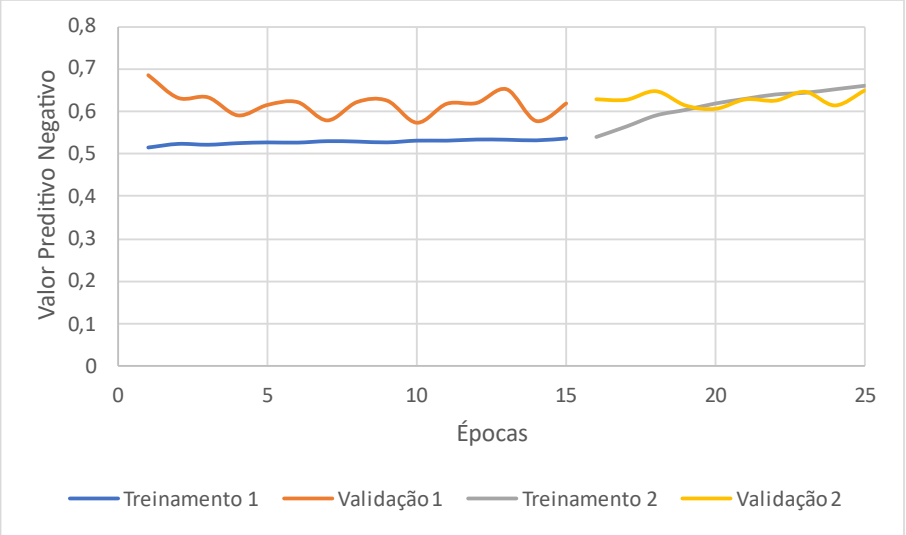
Fonte: Vaz Gabriel (2023)

Figura 14: Sensibilidade da CNN com uma camada convolucional treinável.



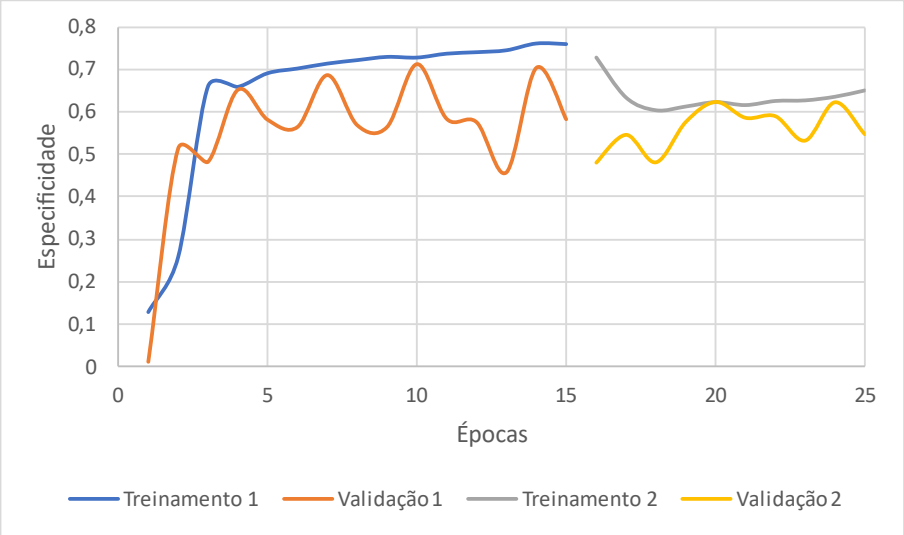
Fonte: Vaz Gabriel (2023)

Figura 15: Valor preditivo negativo da CNN com uma camada convolucional treinável.



Fonte: Vaz Gabriel (2023)

Figura 16: Especificidade da CNN com uma camada convolucional treinável.



Fonte: Vaz Gabriel (2023)

Todos os gráficos apresentam a evolução das métricas de treinamento e validação nas duas etapas de treinamento, neste caso, reduzidas a 25, e alguns pontos importantes devem ser observados.

A primeira observação é o comportamento inicial deste modelo em comparação com o anterior. Ao contrário do primeiro modelo, este iniciou com uma sensibilidade de praticamente perfeita e uma especificidade praticamente nula. Isso ocorre devido à inicialização aleatória dos pesos das redes neurais. A não ser que sejam definidos pesos iniciais para os parâmetros na compilação do modelo, eles são inicializados aleatoriamente e, durante o treinamento, eles são otimizados a fim de obter o menor erro possível. As diferenças obtidas ao final da primeira etapa de treinamento se devem ao fato de que, durante a otimização, o modelo busca por pontos de mínimo no valor do erro, porém, dois treinamentos diferentes, inicializados com pesos diferentes, podem atingir mínimos locais diferentes, ao invés de encontrar o mínimo global.

A segunda observação é que, desconsiderando as diferenças apenas mencionadas, a tendência do comportamento deste modelo é exatamente a mesma do modelo anterior: acurácia praticamente constante, aumento da especificidade simétrica à diminuição da sensibilidade, aumento da precisão síncrono à queda da sensibilidade e valor preditivo negativo constante. Novamente, esse comportamento indica que o modelo começou o treinamento mais sensível à classe positiva

(majoritária) e, durante o treinamento, a sensibilidade diminuiu, aumentando a taxa de verdadeiros negativos, porém, ao longo desse processo, ele também aumentou a taxa de falsos negativos.

A terceira observação é que, apesar da função de retorno para ajuste da taxa de aprendizado estar habilitada, ela não atuou na primeira etapa do treinamento. Por outro lado, a segunda função de retorno, que monitora exclusivamente a taxa de verdadeiros negativos no conjunto de validação, detectou uma estagnação da evolução dessa métrica e interrompeu o treinamento ao término da 15ª época, de forma a evitar um processamento desnecessário (como foi visto no primeiro modelo, não há evolução significativa nas épocas finais de treinamento) e uma eventual degradação da métrica monitorada.

A quarta observação diz respeito à disparidade das métricas dos conjuntos de treinamento e teste já na primeira etapa do treinamento. No entanto, neste caso, o conjunto que está se destacando é o de validação, indicando uma possível contaminação no conjunto de treinamento, uma vez que, com as funções de retorno, há uma tendência de as métricas de validação influenciarem mais fortemente o processo de treinamento. Contudo, na segunda etapa de treinamento, com a habilitação da última camada convolucional da rede VGG-16 para treinamento, nota-se uma correção desse efeito: enquanto as métricas de validação permanecem constantes, as métricas de treinamento retornam para os mesmo níveis, indicando uma melhor capacidade de generalização, tanto no conjunto de treinamento quanto de validação. Isso fica mais evidente ao analisar os resultados obtidos com o conjunto de teste, apresentados na Tabela 7.

Tabela 7: Métricas do conjunto de teste para o segundo modelo treinado.

Etapa	**Acurácia**	**Precisão**	**Sensibilidade**	**Valor preditivo negativo**	**Especificidade**
1	0,67	0,64	0,31	0,67	0,89
2	0,68	0,62	0,41	0,69	0,84

Fonte: Vaz Gabriel (2023)

A evolução dessas métricas, entre a primeira e a segunda etapa de treinamento, que a utilização das funções de retorno com monitoramento de

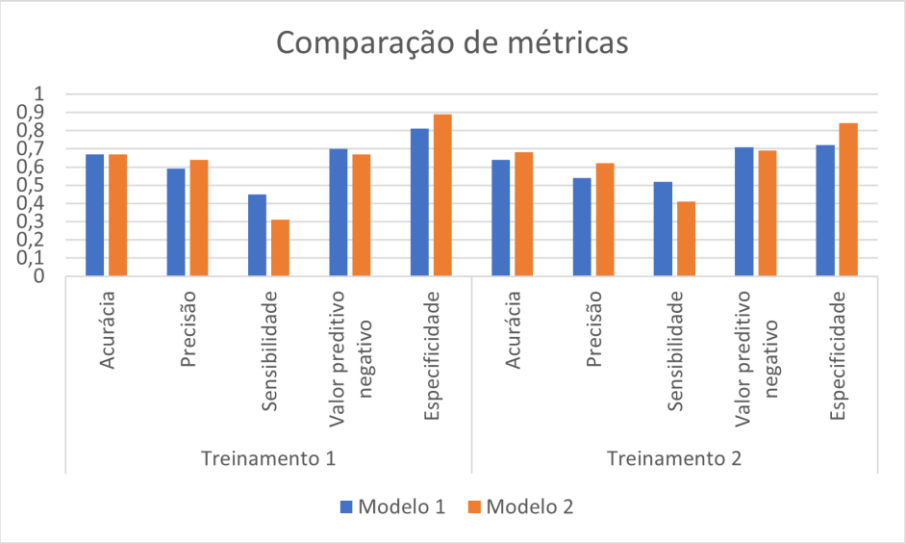
verdadeiros positivos teve uma forte influência nos resultados da especificidade. Porém, essa melhoria se deu em detrimento da sensibilidade, de forma que, ao término da primeira etapa de treinamento, somente 31% dos dados pertencentes à classe positiva eram classificados corretamente. Esse aspecto melhorou após a segunda etapa de treinamento, na qual, como visto anteriormente nos gráficos, diminuiu-se a disparidade entre as métricas de treinamento e validação. Não se pode dizer que o problema foi resolvido, afinal, a sensibilidade ainda permaneceu muito abaixo das demais métricas, contudo, deve-se notar que houve um aumento de 10% na sensibilidade do modelo com uma diminuição de apenas 5% na especificidade.

Considerando que a métrica principal adotada para essa pesquisa é a especificidade, que está relacionada com a capacidade da rede neural em detectar a classe negativa, 84% é um valor compatível com as métricas obtidas em aplicações de visão computacional encontradas na literatura (SEYYED-KALANTARI, LIU, *et al.*, 2020).

5.3 Comparação dos resultados

Os resultados obtidos pelos modelos no conjunto de teste foram compilados e estão apresentados na Figura 17 para facilitar a análise comparativa.

Figura 17: Comparação das métricas no conjunto de teste.



Fonte: Vaz Gabriel (2023)

Na Figura 17, é possível notar que as modificações feitas entre o primeiro e o segundo modelo surtiram efeitos positivos no que diz respeito à classificação da

classe negativa, afinal, tanto na primeira etapa de treinamento quanto na segunda, o segundo modelo apresentou uma especificidade mais alta do que o primeiro, o que refletiu também no nível de precisão dos modelos, uma vez que, quanto maior for o número de verdadeiros negativos, menor será o número de falsos positivos. No entanto, o gráfico também mostra que a melhora da especificidade implicou numa piora da sensibilidade do modelo à classe positiva.

Conforme dito anteriormente, esta é uma aplicação na qual é essencial classificar corretamente a classe negativa e preferível classificar corretamente a classe positiva. Segundo esse critério, pode-se dizer que o segundo modelo atingiu o objetivo, mas há uma margem considerável para se melhorar ainda mais a otimização dessa rede neural para detecção de doenças toráxicas em imagens de raio-x em pesquisas futuras.

6 DISCUSSÃO

Este capítulo tem como objetivo evidenciar os resultados obtidos durante a pesquisa no que diz respeito ao treinamento de redes neurais convolucionais dedicadas a visão computacional na área médica, principalmente no momento atual de evolução tecnológica que permite cada vez mais o compartilhamento de grandes quantidades de dados (*big data*), algo que é essencial para treinamentos de redes neurais e, em especial, otimização de modelos de *deep learning*.

Tais aplicações vêm sendo aprimoradas constantemente e, com os novos recursos, seja com a disponibilidade de *hardware* dedicado para o tratamento de grandes conjuntos de dados ou com a possibilidade de alocação de recursos remotos em nuvem para a mesma finalidade, a possibilidade de estudar novas técnicas na área de aprendizado de máquina está cada vez mais próxima dos pesquisadores, afinal, enquanto as grandes empresas de tecnologia começaram a trabalhar há décadas com os primeiros modelos de inteligência artificial, os acadêmicos só começaram a entrar nessa área há poucos anos.

Com essa “democratização” das técnicas e recursos para de aprendizado de máquina, mostrou-se possível, com um computador relativamente básico (sem múltiplas GPUs, por exemplo), fazer um treinamento de uma rede neural capaz de classificar corretamente 84% das imagens que apresentavam sinais de doenças com um banco de dados de 112120 imagens em 8 horas, o que é um tempo muito pequeno, se for considerado que a rede VGG-16, dadas as devidas proporções e diferenças, precisou ser treinada durante mais de duas semanas.

Outro aspecto que é importante salientar é o equívoco do pensamento popular, que supõe que os sistemas de inteligência artificial (em especial as redes neurais e o *deep learning*) vão eliminar postos de trabalho. O resultado desta pesquisa, embora promissor, mostra que a figura humana é imprescindível, afinal, o modelo treinado classificou 16% dos pacientes enfermos como saudáveis, o que, numa situação real, seria desastroso. Logo, a inteligência artificial é um complemento ao trabalho humano e não um substituto.

7 CONCLUSÃO

Os sistemas computacionais têm desempenhado um papel importante na sociedade desde que os primeiros softwares foram desenvolvidos. No princípio, os programadores desenvolviam aplicações dedicadas a tarefas simples, com o objetivo de poupar os humanos de tarefas complexas e repetitivas, como a geração das folhas de pagamento para todos os funcionários de uma empresa. Esses tipos de aplicações se baseavam em dados exclusivamente estruturados, a partir de bancos de dados com campos e relações de tabelas bem estabelecidas, e despendiam pouco poder computacional.

Com o passar do tempo, no entanto, cada vez mais pessoas passaram a ter acesso a novas formas de geração de dados, de forma que o modelo de computação que havia funcionado por anos se tornou incapaz de lidar com a nova realidade, seja pelo volume de informações ou pelo formato dos novos dados. Nesse contexto, juntamente com a evolução tecnológica que permitiu a criação de *hardwares* com maior poder computacional, surgiram os sistemas de aprendizado de máquina, mais especificamente o *deep learning*, que não só conseguem manipular o *big data*, como também dependem dele para possibilitar o seu funcionamento.

Com este novo paradigma computacional, novos horizontes foram e continuam sendo descobertos, pois, uma vez que o comportamento dos sistemas de aprendizado de máquina não é regido por algoritmos rígidos e imutáveis, os computadores têm uma “liberdade” para aprender novos padrões que é limitada somente pela quantidade de dados disponíveis e pelos limites matemáticos envolvidos no aprendizado. Tal versatilidade se traduz na multidisciplinariedade dessa área da computação, ou seja, qualquer que seja a área do conhecimento na qual se deseje utilizar modelos de *deep learning*, se houver dados suficientes e adequados disponíveis, a proposta é promissora.

Justamente por essa razão, a área médica tem sido um dos principais focos dos pesquisadores de aprendizado de máquina, principalmente da área de visão computacional, desde os seus primórdios, seja para aplicações clássicas, como a deteção de melanomas a partir de fotografias, ou para aplicações mais recentes, como o diagnóstico de Covid-19 a partir de radiografias dos pulmões.

A relevância do *deep learning* em sistemas de diagnósticos médicos foi o que motivou esta pesquisa e, tendo em mente a necessidade de dados para o desenvolvimento de um modelo de rede neural profunda, buscou-se um banco de imagens significativo para esta finalidade, o que levou ao ChestX-ray14, que, além de atender ao requisito de quantidade de dados (112120 imagens), também apresenta grande variabilidade, dado que são catalogadas quatorze doenças num único conjunto de dados.

No entanto, a variabilidade se provou uma faca de dois gumes: ao mesmo tempo que favorece o treinamento da rede neural no sentido que é mais difícil memorizar (*overfitting*) dados variados, também prejudica o treinamento se houver desbalanceamento entre as classes presentes no banco de dados (efeito este que foi observado e discutido nesta pesquisa).

Por outro lado, deve-se ressaltar que existem técnicas que “devolvem” parcialmente o controle do treinamento da rede neural para o programador, de forma que seja possível guiar, por meio de funções de monitoramento em tempo real, a otimização do modelo. Tais técnicas foram exploradas e confirmadas durante o desenvolvimento desta pesquisa

Também é relevante a variedade de métricas de avaliação das redes neurais. Muitas vezes, o desempenho de uma um modelo de aprendizado de máquina é avaliado com base em variáveis genéricas (acurácia) ou abstratas (perda). Esta pesquisa demonstrou que esse tipo de abordagem, dependendo da aplicação, é inadequada, pois, embora resultados positivos tenham sido obtidos em quase todas as métricas analisadas, com exceção da sensibilidade, pelas razões discutidas anteriormente, a métrica mais importante para esta aplicação (especificidade) foi a que atingiu maiores níveis em todos os testes, o que não dispensa novas pesquisas para melhorar o modelo como um todo.

Ainda com relação às métricas de avaliação, pode-se notar que a melhor métrica não foi capaz de atingir 100% de classificações corretas. Por melhor que sejam as redes neurais atuais, esse valor ainda é utópico, portanto, deve-se ressaltar que a introdução do *deep learning* nas diversas áreas nas quais ele pode ser aplicado não implica, necessariamente, na substituição de profissionais humanos,

principalmente em áreas como a médica, na qual vidas dependem dos resultados dos exames.

Um último ponto a ser abordado diz respeito à disponibilidade de infraestruturas para o desenvolvimento de novas técnicas de aprendizado de máquina, seja com equipamentos dedicados ou com recursos de computação em nuvem. Atualmente, os recursos que até alguns anos atrás eram restritos às grandes empresas de tecnologia estão ao alcance de pesquisadores acadêmicos, que podem contribuir com a evolução dessa área do conhecimento, cujo potencial máximo ainda é desconhecido.

8 TRABALHOS FUTUROS

Como trabalhos futuros, serão propostas melhorias deste modelo, afinal, os resultados mostraram que há margem para otimizar e aumentar as métricas de avaliação da rede neural. Com essa perspectiva, algumas abordagens podem ser feitas: *data augmentation* para aumentar a quantidade de dados de treinamento, principalmente aqueles pouco representativos (classes com menos de 10000 imagens); busca de bancos de dados complementares para aumentar o conjunto de treinamento; estudo do dimensionamento de hiper parâmetros que possam levar a rede ao menor erro possível (mínimo global); estudo de novas funções de retorno, que monitorem o treinamento em tempo real e auxiliem na otimização do modelo.

Outra possibilidade é a mudança do objetivo da rede neural. Uma vez que seja encontrada uma técnica de *data augmentation* adequada, seria possível criar uma rede neural para classificação *multilabel*, ou seja, que, além de identificar a presença ou ausência de doenças, seja capaz de identificar a doença propriamente dita.

REFERÊNCIAS

ACESSA.COM. **O que é Big Data?** [S.l.], p. https://www.acessa.com/tecnologia/arquivo/artigo/2018/06/18-que-big-data/. 2018.

AGGARWAL, R. et al. Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. **NPJ digital medicine**, v. 4, p. 1-23, 2021.

AL., L. E. Backpropagation Applied to Handwritten Zip Code Recognition. **Neural Computation**, p. 541–551, 1989.

ALBER, M. et al. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. **NPJ digital medicine**, v. 2, p. 1-11, 2019.

ANTUNES, E. et al. **Gestão da Tecnologia Biomédica, Tecnovigilância e Engenharia Clínica**. 1ª. ed. [S.l.]: Acodess, 2002.

ARCAS, B. A. Y. Art in the Age of Machine Intelligence. **MDPI Arts**, 2017.

BAIA, A. F.; CASTRO, A. R. G. **A Competitive Structure of Convolutional Autoencoder Networks for Electrocardiogram Signals Classification**. Anais do XV Encontro Nacional de Inteligência Artificial e Computacional. [S.l.]: SBC. 2018. p. 538-549.

BAKER, J. et al. Research Developments and Directions in Speech Recognition and Understanding, Part 1. **IEEE Signal Processing Magazine**, 2009. 75-80.

BARNES, T. J. Big data, little history. **Sage Journals** , 10 Dezembro 2013. 297-302.

BATISTIč, S. A. P. V. D. L. History, evolution and future of big data and analytics: a bibliometric analysis of its relationship to performance in organizations. **British Journal of Management**, 2019. 229-251.

BATTY, M. Big Data and the City. **Bult Environment**, v. 42, n. 1, p. 321-337, Outubro 2016. ISSN 3.

BEAM, A. L. . A. I. S. K. Big data and machine learning in health care. **Jama**, 2018. 1317-1318.

BEHNKE, S. (. Hierarchical Neural Networks for Image Interpretation. **Lecture Notes in Computer Science.**, 2003.

BENGIO, Y. **Artificial Neural Networks and their Application to Speech/Sequence Recognition**. McGill University Ph.D. thesis.. [S.l.], p. https://www.researchgate.net/publication/41229141. 1991.

BENGIO, Y. "Learning Deep Architectures for AI". **Foundations and Trends in Machine Learnin**, v. 2, n. 1, p. 1–127. , 2009.

BENGIO, Y. Practical recommendations for gradient-based training of deep architectures, 2012.

BENGIO, Y. et al. Greedy layer-wise training of deep networks. **Advances in neural information processing systems.**, p. 153–160, 2007.

BENGIO, Y. et al. Towards Biologically Plausible Deep Learning, 2015.

BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation Learning: A Review and New Perspectives. **EEE Transactions on Pattern Analysis and Machine Intelligence.** , p. 1798–1828, 2013.

BENGIO, Y.; LECUN, Y.; HINTON, G. Deep Learning. **Nature**, p. 436–444, 2015.

BISHOP, C. M. Pattern Recognition and Machine Learning. **Springer**, 2017.

BISWAS, S. et al. Low-N protein engineering with data-efficient deep learning. **Nature methods**, v. 18, p. 389-396, 2021.

BRADLEY KNOX, W.; STONE, P. TAMER: Training an Agent Manually via Evaluative Reinforcement. **7th IEEE International Conference on Development and Learning**, 2008. 292-297.

BRONZINO, J. Biomedical Engineering: A Historical Perspective. In: ENDERLE, J.; BLANCHARD, S.; BRONZINO, J. **Introduction to Biomedical Engineering**. 2ª. ed. [S.l.]: Academic Press, 2005. Cap. 1, p. 1-29.

BUESING, L. et al. Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons. **PLOS Computational Biology**, 2011.

BUGHIN, J. Reaping the benefits of big data in telecom. **Journal of Big Data** , 2016. 1-17.

BUREAU INTERNATIONAL DES POIDS ET MESURES. **Resolutions of the General Conference on Weights and Measures (27th meeting)**. Bureau International des Poids et Mesures. Sèvres, p. 22. 2022.

BURNS, M. Blippar Demonstrates New Real-Time Augmented Reality App. **Techcrunch**, 2015. Disponivel em: <https://techcrunch.com/2015/12/08/blippar-demonstrates-new-real-time-augmented-reality-app/>. Acesso em: 21 Setembro 2021.

CARDENAS, A. A. . P. K. M. A. S. P. R. Big data analytics for security. **IEEE Security & Privacy**, 2013. 74-76.

CASH, S.; YUSTE, R. Linear summation of excitatory inputs by CA1 pyramidal neurons. **Neuron**, 1999. 383–394.

CHAO, H. et al. Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography. **Nature communications**, v. 12, p. 1-10, 2021.

CHEN, C.-L. et al. An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning. **Nature communications**, v. 12, p. 1-13, 2021.

CHICCO, D.; SADOWSKI, P.; BALDI, P. Deep Autoencoder Neural Networks for Gene Ontology Annotation Predictions. **Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics - BCB '14. ACM.**, 2014. 533–540..

CHO, H.-H. et al. Radiomics-guided deep neural networks stratify lung adenocarcinoma prognosis from CT scans. **Communications biology**, v. 4, p. 1-12, 2021.

CHOI, E. et al. Using recurrent neural network models for early detection of heart failure onset. **Journal of the American Medical Informatics Association.**, 2016. 361-370.

CIRESAN, D. C. et al. Flexible, High Performance Convolutional Neural Networks for Image Classification. **International Joint Conference on Artificial Intelligence.**, 2011.

CIRESAN, D. C. et al. Flexible, High Performance Convolutional Neural Networks for Image Classification. **International Joint Conference on Artificial Intelligence**, 2021. https://doi.org/10.5591%2F978-1-57735-516-8%2Fijcai11-210.

CIREşAN, D. et al. Multi-column deep neural network for traffic sign classification". Neural Networks. **IJCNN**, 2011.

CIRESAN, D. et al. Advances in Neural Information Processing Systems , 2012. 2843–2851.

CIRESAN, D. et al. Mitosis Detection in Breast Cancer Histology Images using Deep Neural Networks. **Proceedings MICCAI. Lecture Notes in Computer Science.** , 2013. 11–418.

CIRESAN, D.; MEIER, U.; SCHMIDHUBER, J. "Multi-column deep neural networks for image classification".. **EEE Conference on Computer Vision and Pattern Recognition**, p. 3642–3649. , 2012.

CO-EVOLVING recurrent neurons learn deep memory POMDPs. Proc. **ACM Press**, New York, p. 1795-1802, USA.

COURSERA. Data Augmentation. **Coursera**. Disponivel em: <https://www.coursera.org/lecture/convolutional-neural-networks/data-augmentation-AYzbX>. Acesso em: 22 Setembro 2021.

CUI, L. F. R. Y. A. Q. Y. When big data meets software-defined networking: SDN for big data and big data for SDN. **IEEE network**, 2016. 58-65.

CYBENKO. "Approximations by superpositions of sigmoidal functions". **Mathematics of Control, Signals, and Systems.** , v. 2, n. 4, p. 303–314., 1989.

CZECH, T. Deep learning: the next frontier for money laundering detection. **Global Banking and Finance Review**, 2018.

D. YU, L. D. G. L. A. F. S. **Discriminative pretraining of deep neural networks**. U.S. Patent Filing., 2011.

DAHL, G.; SAINATH, T.; HINTON, G. IMPROVING DEEP NEURAL NETWORKS FOR LVCSR USING RECTIFIED LINEAR UNITS. **ICASSP**, 2013. 06-13.

DAHL, G.; SAINATH, T.; HINTON, G. Improving DNNs for LVCSR using rectified linear units and dropout. **ICASSP**, 2013.

DAILY, J. A. J. P. Predictive maintenance: How big data analysis can improve maintenance. **Supply chain integration challenges in commercial aerospace. Springer, Cham**, 2017. 267-278.

DAS, P. et al. Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. **Nature Biomedical Engineering**, v. 5, p. 613-623, 2021.

DE CARVALHO, A. C. L. F.; FAIRHURST, M. C.; BISSET, D. An integrated Boolean neural network for pattern classification". Pattern Recognition Letters., v. 15, n. 8, p. 807-813, 1994.

DE, S. et al. Predicting the popularity of instagram posts for a lifestyle magazine using deep learning. **2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA).** , 174-177 2017.

DECHTER, R. Learning while searching in constraint-satisfaction problems.. **University of California, Computer Science Department, Cognitive Systems Laboratory**, p. 4-19, 2016.

DENG, L. et al. Recent advances in deep learning for speech research at Microsoft. **IEEE International Conference on Acoustics, Speech and Signal Processing**, p. 8604–8608. , 2013.

DENG, L.; ABDEL-HAMID, O.; YU, D. A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion. **ICASSP**, 2013.

DENG, L.; HASSANEIN, K.; ELMASRY, M. Analysis of correlation structure for a neural predictive model with applications to speech recognition. **Neural Networks**, 1994. 331–339.

DENG, L.; HINTON, G.; KINGSBURY, B. New types of deep neural network learning for speech recognition and related applications: An overview. **Microsoft Research.**, 2017.

DENG, L.; HINTON, G.; KINGSBURY, B. New types of deep neural network learning for speech recognition and related applications: An overview. **ICASSP**, 2017.

DENG, L.; YU, D. "Deep Learning: Methods and Applications"**.. Foundations and Trends in Signal Processing**, v. 7, n. (3–4), p. 1-1999, 2014.

DEOLDIFY: Colorizing and Restoring Old Images and Videos with Deep Learning. **Floydhub**, 2018. Disponivel em: <https://blog.floydhub.com/colorizing-and-restoring-old-images-with-deep-learning/>. Acesso em: 2021 Setembro 21.

DIAZ-AVILES, E. E. A. Towards real-time customer experience prediction for telecommunication operators. **IEEE International Conference on Big Data (Big Data).**, 2015.

DIJKSTRA, M.; LUIJTEN, E. From predictive modelling to machine learning and reverse engineering of colloidal self-assembly. **Nature materials**, v. 20, p. 762-773, 2021.

DODDINGTON, G. et al. The NIST speaker recognition evaluation ± Overview, methodology, systems, results, perspective. **Speech Communication**, 2000. 225–254.

EIJNATTEN, J. V. T. P. A. J. V. Big data for global history: The transformative promise of digital humanities. **BMGN-Low Countries Historical Review**, 2013. 55-77.

ELKAHKY, A. M.; SONG, Y.; HE, X. A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems. **Microsoft Research. Archived**, 2015.

ELMAN, J. L. **Rethinking Innateness:** A Connectionist Perspective on Development. [S.l.]: MIT Press, 1988.

ENDERLE, J. **Introduction to biomedical engineering**. 3ª. ed. [S.l.]: Academic Press, 2012.

FENG, X. Y. et al. The Deep Learning–Based Recommender System “Pubmender” for Choosing a Biomedical Publication Venue: Development and Validation Study. **JRM - Journal of Medical Internet Research** , 2019. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6555124/.

FONSECA, M. et al. **VR4NEUROPAIN:** Interactive Rehabilitation System. Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies - BIODEVICES. Prague: Scitepress. 2019. p. 285-290.

FORSLID, G. et al. Deep Convolutional Neural Networks for Detecting Cellular Changes Due to Malignancy. **EEE International Conference on Computer Vision Workshops (ICCVW)**, 2017. 82-89.

FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. **Biol. Cybern**, v. 36, n. 4, p. 193-202, 1980.

G. E. HINTON. Learning multiple layers of representation Archived. **rends in Cognitive Sciences**, p. 428–434, 2007.

GABRIEL, A. T. et al. Development and clinical application of Vertebral Metrics: using a stereo vision system to assess the spine. **Medical & Biological Engineering & Computing**, 20 Janeiro 2018. 1435-1446.

GARLING, C. Startup Harnesses Supercomputers to Seek Cures. **KQED**, 2015. Disponivel em: <https://www.kqed.org/futureofyou/3461/startup-harnesses-supercomputers-to-seek-cures>. Acesso em: 2021 Setembro 2021.

GERS, F. A.; SCHMIDHUBER, J. LSTM Recurrent Networks Learn Simple Context Free and Context Sensitive Languages. **IEEE Transactions on Neural Networks**, 2001. 1333-1340.

GIBNEY, E. "Google AI algorithm masters ancient game of Go. **Nature**, 2016. 445–446.

GILLICK, D. et al. Multilingual Language Processing from Bytes, 2015.

GOES, P. B. Editor's comments: big data and IS research. **JSTOR** , Setembro 2014.

GOLDBERG, Y.; LEVY, O. word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method, 2014.

GRAVES, A. et al. Biologically Plausible Speech Recognition with LSTM Neural Nets. **1st Intl. Workshop on Biologically Inspired Approaches to Advanced Information Technology, Bio-ADIT.** , Lausanne, Switzerland, p. 175–184, 2003.

GRAVES, A.; FERNÁNDEZ, S.; GOMEZ, F. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. **Proceedings of the International Conference on Machine Learning, ICML** , p. 369–376. , 2006.

GREGORY, B. A Molecule Designed By AI Exhibits 'Druglike' Qualities. **Wired**, 2020. https://www.wired.com/story/molecule-designed-ai-exhibits-druglike-qualities/.

GRIEWANK, A. "Who Invented the Reverse Mode of Differentiation?", p. 389–400, 2012.

GÜÇLÜ, U.; VAN GERVEN, M. A. J. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. **ournal of Neuroscience.** , 2015. 10005–10014.

GUTMANN, M. P. . E. K. M. A. E. R. Big data” in economic history. **The journal of economic history**, 78.1, 03 Abril 2018. 268-299.

HECK, L. et al. Robustness to Telephone Handset Distortion in Speaker Recognition by Discriminative Feature Design. **Speech Communication**, 2000. 181-182.

HINTON, G. E. Learning multiple layers of representation. **Trends in Cognitive Sciences**, v. 11, n. 10, p. 428–434. , 2007.

HINTON, G. E. Deep belief networks. **Scholarpedia**, v. 4, n. 5, p. 5947, 2009.

HINTON, G. E. et al. The wake-sleep algorithm for unsupervised neural networks". Science., v. 258, n. 5214, p. 1158–1161., 1995.

HINTON, G. E. et al. mproving neural networks by preventing co-adaptation of feature detectors, 2012.

HINTON, G. E.; OSINDERO, S.; TEH, Y. W. A Fast Learning Algorithm for Deep Belief Nets. **Neural Computation.**, p. 1527–1554. , 2015.

HINTON, G. et al. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. **EEE Signal Processing Magazine.**, v. 29, n. 6, p. 82-97, 2012.

HOCHREITER, S. et al. **Gradient flow in recurrent nets:** the difficulty of learning long-term dependencies. [S.l.]: [s.n.], 2001.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation.**, v. 9, n. 8, p. 1735–1780. , 1997.

HOF, R. D. Is Artificial Intelligence Finally Coming into Its Own. **MIT Technology Review. Archived from the origina**, 2018.

HOLMLUND, M. E. A. Customer experience management in the age of big data analytics: A strategic framework. **Journal of Business Research** , 2020. 356-365.

HORNIK, K. Approximation Capabilities of Multilayer Feedforward Networks. **Neural Networks**, v. 4, n. 2, p. 251-257, 1991.

HU, J. et al. Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning".. **IEEE Transactions on Vehicular Technology.**, p. 14413–14423., 2020.

HUANG, L. et al. Machine learning of serum metabolic patterns encodes early-stage lung adenocarcinoma. **Nature communications**, v. 11, p. 1-11, 2020.

IGOR AIZENBERG, N. N. A. J. P. L. V. Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications.. **Springer Science & Business Media.**

INSPER INSTITUTO DE ENSINO E PESQUISA. O que mudou em 10 anos na lista das empresas mais valiosas do mundo, 2022. Disponivel em: <https://www.insper.edu.br/noticias/o-que-mudou-em-10-anos-na-lista-das-empresas-mais-valiosas-do-mundo/>. Acesso em: 5 Janeiro 2023.

IVAKHNENKO, A. Polynomial theory of complex systems. **IEEE Transactions on Systems, Man and Cybernetics. SMC-**, p. 364-378, 1971.

IVAKHNENKO, A. G.; LAPA, V. G. Cybernetics and Forecasting Techniques. **American Elsevier Publishing Co**, 1967.

JAGADISH, H. V. . E. A. Big data and its technical challenges. **Communications of the ACM**, 2014. 86-94.

JIFA, G. A. Z. L. Data, DIKW, Big data and Data science. **Procedia computer science** , 2014. 814-821.

JOHNSON, J. S. . S. B. F. A. H. S. L. Big data facilitation, utilization, and monetization: Exploring the 3Vs in a new product development process. **Journal of Product Innovation Management** , 2017. 640-658.

KANAVATI, F. et al. Weakly-supervised learning for lung carcinoma classification using deep learning. **Scientific reports**, v. 10, p. 1-11, 2020.

KEROUAC, J. **Jack Kerouac: Um dia hei de renascer numa grande.** [S.l.], p. https://www.pensador.com/frase/NTM0Njkw/. 2005.

KIM, Y. et al. Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. **Scientific reports**, v. 11, p. 1-11, 2021.

KING, N. M. R. A. J. H. hree paradoxes of big data. **Stan. L. Rev. Online**, 2013. 66.

KIROS, R.; SALAKHUTDINOV, R.; ZEMEL, R. S. Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models, 2014.

KLEANTHOUS, C.; CHATZIS, S. Gated Mixture Variational Autoencoders for Value Added Tax audit case selection. **nowledge-Based Systems**, 2020.

KNIGHT, W. DARPA is funding projects that will try to open up AI's black boxes. **MIT Technology Review. Archived**, 2017.

KOBIELUS, J. GPUs Continue to Dominate the AI Accelerator Market for Now. **InformationWeek**, 2019. Disponivel em: <https://www.informationweek.com/ai-or-machine-learning/gpus-continue-to-dominate-the-ai-accelerator-market-for-now>. Acesso em: 22 Setembro 2021.

KOSELEVA, N. A. G. R. Big data in building energy efficiency: understanding of big data and main challenges. **Procedia Engineering**, 2017. 544-549.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. "ImageNet Classification with Deep Convolutional Neural Networks". **NIPS 2012: Neural Information Processing Systems, Lake Tahoe, Nevada.** , p. 05-24, 2012.

L. GU, D. Z. P. L. A. S. G. Cost Minimization for Big Data Processing in Geo-Distributed Data Centers. **EEE Transactions on Emerging Topics in Computing**, Setembro 2013. 314-323.

L'HEUREUX, A. E. A. Machine learning with big data: Challenges and approaches. **IEEE**, 2017. 7776-7797.

LABORATORY, A. R. Army researchers develop new algorithms to train robots. **EurekAlert**, 2018. Disponivel em: <https://www.eurekalert.org/news-releases/754849>. Acesso em: 2021 Setembro 21.

LABRINIDIS, A. A. H. V. J. Challenges and opportunities with big data. **Proceedings of the VLDB Endowment** , 2012. 2032-2033.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015.

LI, D. Keynote talk: 'Achievements and Challenges of Deep Learning - From Speech Analysis and Recognition To Language and Multimodal Processing, 2014. https://en.wikipedia.org/wiki/Deep_learning#:~:text=).%20%22Keynote%20talk%3A%20%27Achievements%20and%20Challenges%20of%20Deep%20Learning%20-%20From%20Speech%20Analysis%20and%20Recognition%20To%20Language%20and%20Multimodal%20Processing%27.

LI, J. E. A. Big data in product lifecycle management. **The International Journal of Advanced Manufacturing Technology**, 2015. 667-684.

LI, X.; WU, X. Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition, 2014.

LINNAINMAA, S. The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors. Master's Thesis (in Finnish). **Univ. Helsinki**, p. 6-7, 1970.

LITJENS, G. et al. A survey on deep learning in medical image analysis. **Medical Image Analysis.** , 2017. 60-88.

LOHR, S.". How big data became so big. **New York Times**, 2012. Disponivel em: <http://euler.mat.ufrgs.br/~viali/estatistica/estatistica/relacionados/13_Big_Data_So_Big.pdf>. Acesso em: 20 Setembro 2021.

LU, Z. . P. H. . W. F. . H. Z. . &. W. L. The Expressive Power of Neural Networks: A View from the Width Archived. **Wayback Machine. Neural Information Processing Systems**, 2017.

MANNING, P. **Big data in history**. [S.l.]: Springer, 2013.

MARBLESTONE, A. H.; WAYNE, G.; KORDING, K. P. Toward an Integration of Deep Learning and Neuroscience. **Frontiers in Computational Neuroscience.**, p. 10-94.

MARCUS, G. Deep Learning: a Revolution in Artificial Intelligence. **he New Yorker. Archived**, 2017.

MARCUS, G. "In defense of skepticism about deep learning. **ary Marcus. Archived**, 2018.

MARX, V. The big challenges of big data. **Nature**, 2013. 255-260.

MCCANEY, K. Talk to the Algorithms: AI Becomes a Faster Learner. **Governmentciomedia**, 2018. Disponivel em: <https://governmentciomedia.com/talk-algorithms-ai-becomes-faster-learner>. Acesso em: 2021 Setembro 21.

METZ, C. Facebook's 'Deep Learning' Guru Reveals the Future of AI. **Wired. Archived**, 2014.

METZ, C.. "A.I. Researchers Leave Elon Musk Lab to Begin Robotics Start-Up". **The New York Times**, 2019. Disponivel em: <https://www.nytimes.com/2017/11/06/technology/artificial-intelligence-start-up.html>. Acesso em: 21 Setembro 2021.

MIKOLOV, T.; AL., E. Recurrent neural network based language model. **ISCA**, 2010. 1045–1048.

MIT TECHNOLOGY REVIEW. **MIT Technology Review**, 2016. Disponivel em: <http://www.technologyreview.com/news/546066/googles-ai-masters-the-game-of-go-a-decade-earlier-than-expected/>. Acesso em: 21 Setembro 2021.

MOEN, E. et al. Deep learning for cellular image analysis. **Nature methods**, v. 16, p. 1233-1246, 2019.

MOREL, D.; SINGH, C.; LEVY, W. B. Linearization of excitatory synaptic integration at no extra cost".. **Journal of Computational Neuroscience.**, 2010. 173–188.

MORGAN, N. et al. Hybrid neural network/hidden markov model systems for continuous speech recognition. **International Journal of Pattern Recognition and Artificial Intelligence**, 1983. 899–916..

MURPHY, K. P. Machine Learning: A Probabilistic Perspective. **MIT Press.**, 2012.

NATIONAL CENTER FOR ADVANCING TRANSLATIONAL SCIENCES. Final Subchallenge Leaderboard. **National Center for Advancing TransLational Sciences**, 2014. Disponivel em: <https://tripod.nih.gov/tox21/challenge/leaderboard.jsp>. Acesso em: 22 Setembro 2021.

NOGARE, D. Performance de Machine Learning - Matriz de Confusão - Diego Nogare. **Diego Nogare - Inteligência Artificial e Machine Learning**, 2020. Disponivel em: <https://diegonogare.net/2020/04/performance-de-machine-learning-matriz-de-confusao/>. Acesso em: 10 Janeiro 2023.

OAKDEN-RAYNER, L. et al. Precision radiology: predicting longevity using feature engineering and deep learning methods in a radiomics framework. **Scientific reports**, v. 7, p. 1-13, 2017.

OHM, P. The underwhelming benefits of big data. **U. Pa. L. Rev. PENNumbra** , 2012. 339.

OLSHAUSEN, B. A. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. **Nature** , 1996.

OLSHAUSEN, B.; FIELD, D. Sparse coding of sensory inputs".. **Current Opinion in Neurobiolog**, 2004. 481–487..

O'MALLEY, M. Doing what works: Governing in the age of big data. **Public Administration Review** , 2014. 555-556.

ORACLE. **O que é Big Data?** [S.l.], p. https://www.oracle.com/br/big-data/what-is-big-data/. 2021.

QIU, J. E. A. A survey of machine learning for big data processing. **EURASIP Journal on Advances in Signal Processing**, 2016. 1-16.

RAGUSEO, E. Big data technologies: An empirical investigation on their adoption, benefits and risks for companies." **International Journal of Information Management**, 2018. 187-195.

RICHARD SOCHER, A. P. J. Y. W. J. C. Recursive Deep Models for Semantic Compositionality, 2013.

ROBINSON, T. Several Improvements to a Recurrent Error Propagation Network Phone Recognition System. **Cambridge University Engineering Department Technical Report. CUED/F-INFENG/TR82**, 1991.

ROBINSON, T. A real-time recurrent error propagation network word recognition system. **ICASSP**, 1992. 617–620.

ROHINI, G. Everything you need to know about VGG16 | by Great Learning | Medium. **Medium**, 2021. Disponivel em: <https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918>. Acesso em: 10 Janeiro 2023.

ROLNICK, D.; TEGMARK, M. The power of deeper networks for expressing natural functions. **International Conference on Learning Representations. ICLR** , 2018.

RONDON, I. F. et al. **Engenharia Biomédica 1**. 1ª. ed. [S.l.]: Independently published, 2018.

S. HOCHREITER. Untersuchungen zu dynamischen neuronalen Netzen Archived. **nstitut f. Informatik, Technische Univ. Munich. Advisor: J. Schmidhuber**, 2015.

SAK, H. et al. Google voice search: faster and more accurate, 2015.

SAK, H.; SENIOR, A.; BEAUFAYS, F. Long Short-Term Memory recurrent neural network architectures for large scale acoustic modeling, 2014.

SANTIAGO FERNANDEZ, A. G. A. J. S. An application of recurrent neural networks to discriminative keyword spotting Archived. **Wayback Machine. Proceedings of ICANN** , p. 220–229., 2007.

SATHYANARAYANA, A. Sleep Quality Prediction From Wearable Data Using Deep Learning. **JMIR mHealth and uHealth**, 2016.

SCHMIDHUBER, J. Deep Learning. **Scholarpedia**, v. 10, n. 11, 2015.

SCHMIDHUBER, J. Deep Learning in Neural Networks: An Overview".. **Neural Networks.** , v. 85-117, 2015.

SEYYED-KALANTARI, L. et al. **CheXclusion:** Fairness gaps in deep chest X-ray classifiers. BIOCOMPUTING 2021: proceedings of the Pacific symposium. [S.l.]: World Scientific. 2020. p. 232-243.

SHIGEKI, S.. Human Behavior and Another Kind in Consciousness: Emerging Research and Opportunities: Emerging Research and Opportunities. **IGII GLOBAL**, Abril 2019.

SHRAGER, J.; JOHNSON, M. Dynamic plasticity influences the emergence of function in a simple cortical array. **Neural Networks**, 1996. 1119–1129.

SHRAGER, J.; JOHNSON, M. Dynamic plasticity influences the emergence of function in a simple cortical array. **Neural Networks.**, 1996. 1119–1129.

SILVER, D. et al. Mastering the game of Go with deep neural networks and tree search. **Nature**, 2016. 484–489..

SILVER, D. et al. Mastering the game of Go with deep neural networks and tree search. **Nature**, 2016. 484-489.

SIMONYAN, K.; ZISSERMAN, A. **Very deep convolutional networks for large-scale image recognition**. 3rd International Conference on Learning Representations (ICLR 2015). San Diego: Computational and Biological Learning Society. 2015. p. 1-14.

SINANC, S. S. A. D. Big data: A review. **International Conference on Collaboration Technologies and Systems (CTS)**, 2013. 42-47.

SINGH, P.; SAHA, G.; SAHIDULLAH, M. Non-linear frequency warping using constant-Q transformation for speech emotion recognition. **International Conference on Computer Communication and Informatics (ICCCI).**, p. 1-4, 2021.

SMITH, G. W.; LEYMARIE, F. F. The Machine as Artist: An Introduction. **MDPI - Arts**, 2017.

SOCHER, R.; MANNING, C. Deep Learning for NLP, 2014.

SOCHER, R.; MANNING, C. Deep Learning for NLP , 2014.

SONODA, S.; MURATA, N. eural network with unbounded activation functions is universal approximator. **Applied and Computational Harmonic Analysis**, v. 43, n. 2, p. 233-268, 2017.

SPIESS, J. E. A. Using big data to improve customer experience and business performance. **Bell labs technical journal** , 2014. 3-17.

SRIVASTAV, S. Artificial Intelligence, Machine Learning, and Deep Learning. What's the Real Difference? **Start it up**, 2020. Disponivel em: <https://medium.com/swlh/artificial-intelligence-machine-learning-and-deep-learning-whats-the-real-difference-94fe7e528097>. Acesso em: 22 Setembro 2021.

SUTSKEVER, L.; VINYALS, O.; LE, Q. Sequence to Sequence Learning with Neural Networks, 2014.

SZE, V. et al. Efficient Processing of Deep Neural Networks. **A Tutorial and Survey**, 2017.

SZEGEDY, C.; TOSHEV, A.; ERHAN, D. Deep neural networks for object detection. **Advances in Neural Information Processing Systems**, 2013. 2553–2561.

TAPPERT, C. C. "Who Is the Father of Deep Learning? **International Conference on Computational Science and Computational Intelligence (CSCI). IEEE**, p. 343-348, 2019.

TAURION, C. **Big data**. [S.l.]: [s.n.], 2013.

TECHCRUNCH. "Google's AlphaGo AI wins three-match series against the world's best Go player". **TechCrunch**, 2017. Disponivel em: <https://techcrunch.com/2017/05/24/alphago-beats-planets-best-human-go-player-ke-jie/amp/>. Acesso em: 2021 Setembro 21.

TESTOLIN, A.; STOIANOV, I.; ZORZI, M. "Letter perception emerges from unsupervised deep learning and recycling of natural image features. **Nature Human Behaviour.**, 2017. 657–664..

TESTOLIN, A.; ZORZI, M. Probabilistic Models and Generative Neural Networks: Towards an Unified Framework for Modeling Normal and Impaired Neurocognitive Functions. **Frontiers in Computation Neuroscience**, 2016.

THE GLOBE AND MAIL. Toronto startup has a faster way to discover effective medicines. **theglobeandmail**, 2015. Disponivel em: <https://www.theglobeandmail.com/report-on-business/small-business/startups/toronto-startup-has-a-faster-way-to-discover-effective-medicines/article25660419/>. Acesso em: 22 Setembro 2021.

TING QIN, E. A. Continuous CMAC-QRLS and its systolic array. **Neural Processing Letters**, 2005. 1-16.

TOLSTOY, N. M. A. Big Data, Fast Data and Data Lake Concepts. **Procedia Computer Science** , 20 Setembro 2021. 300-305.

TRIERWEILER, M. Evaluation the use of big data analytics to facilitate compliance and fraud prevention: an empirical study about usefulness and usage of big data analytics to prevent occupational fraud in German speaking. **Michaela Trierweiler. Diss. Universität Linz,** 2019.

VAN DEN OORD, A.; DIELEMAN, S.; SCHRAUWEN, B. Deep content-based music recommendation. **Advances in Neural Information Processing System**, 2013. https://proceedings.neurips.cc/paper/2013/file/b3ba8f1bee1238a2f37603d90b58898d-Paper.pdf.

VAN DER LAAK, J.; LITJENS, G.; CIOMPI, F. Deep learning in histopathology: the path to the clinic. **Nature medicine**, v. 27, p. 775-784, 2021.

VEGA, J. J. C.; ORTEGA, J. F. C.; JOYANES-AGUILAR, L. Conociendo Big Data. **Revista Ingenieria**, Tunja, juno 2015. 2-3.

VIEBKE, A. et al. CHAOS: a parallelization scheme for training convolutional neural networks on Intel Xeon Phi. **The Journal of Supercomputing**, 2019. 197–227..

WAIBEL, A. et al. Phoneme recognition using time-delay neural networks. **IEEE Transactions on Acoustics, Speech, and Signal Processing.**, 1989. 328-339.

WANG, G. et al. A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest X-ray images. **Nature biomedical engineering**, v. 5, p. 509-521, 2021.

WANG, X. et al. **ChestX-ray8:** Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE. 2017. p. 3462-3471.

WANG, Y. L. K. A. T. A. B. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. **Technological Forecasting and Social Change** , 2018. 3-13.

WENG, J.; AHUJA, N.; HUANG, T. **Cresceptron:** a self-organizing neural network which grows adaptively. IJCNN International Joint Conference on Neural Networks. [S.l.]: IEEE. 1992. p. 576-581.

WERBOS, P. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. **Harvard University**, 1974.

WERBOS, P. Applications of advances in nonlinear sensitivity analysis. **System modeling and optimization. Springer**, p. 762–770., 1982.

WU, C. R. B. A. K. R. Big data analytics= machine learning+ cloud computing. **arXiv**, 2016.

WU, Y. E. A. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translatio, 2016.

YAMINS, D. L. K.; DICARLO, J. J. Using goal-driven deep learning models to understand sensory cortex. **Nature Neuroscience.** , 2016. 356–365.

YAN, J. E. A. Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance. **IEEE**, 2017. 23484-23491.

YU, D.; DENG, L. Roles of Pre-Training and Fine-Tuning in Context-Dependent DBN-HMMs for Real-World Speech Recognition. **Workshop on Deep Learning and Unsupervised Feature Learning.** , 2010. https://www.microsoft.com/en-us/research/publication/roles-of-pre-training-and-fine-tuning-in-context-dependent-dbn-hmms-for-real-world-speech-recognition/.

YU, D.; DENG, L. Automatic Speech Recognition: A Deep Learning Approach. **sPRINGER**, 2014. https://books.google.com/books?id=rUBTBQAAQBAJ.

ZEN, H.; SAK, H. Unidirectional Long Short-Term Memory Recurrent Neural Network with Recurrent Output Layer for Low-Latency Speech Synthesis. **ICASSP**, p. 4470–4474, 2015.

ZHAN, Y. E. A. Unlocking the power of big data in new product development. **Annals of Operations Research**, 2018. 577-595.

ZHANG, A. et al. Shifting machine learning for healthcare from development to deployment and from models to data. **Nature Biomedical Engineering**, p. 1-16, 2022.

ZHAVORONKOV, A. eep learning enables rapid identification of potent DDR1 kinase inhibitors. **Nature Biotechnology**, 2019. 1038–1040..

ZHONG, S.-H.; LIU, Y.; LIU, Y. Bilinear Deep Learning for Image Classification. **roceedings of the 19th ACM International Conference on Multimedia. MM**, New York, 2011.

ZHOU, L. E. A. Machine learning on big data: Opportunities and challenges. **Neurocomputin**, 2017. 350-361.

ZIBIN ZHENG, J. Z. M. R. L. Service-Generated Big Data and Big Data-as-a-Service: An Overview. **EEE International Congress on Big Data**, 2013. 403-410.

ZIMPHER, N. L. B. Building a smarter university: Big data, innovation, and analytics. **Suny Press**, 2014.

ZOLTÁN TÜSKE, P. G. H. N. Acoustic Modeling with Deep Neural Networks Using Raw Time Signal for LVCSR. **ResearchGate**, 2014. Disponivel em: <https://www.researchgate.net/publication/266030526_Acoustic_Modeling_with_Deep_Neural_Networks_Using_Raw_Time_Signal_for_LVCSR>. Acesso em: 21 sETEMBRO 2021.

ZORZI, M.; TESTOLIN, A. (. F. 2. An emergentist perspective on the origin of number sense. **Phil. Trans. R. Soc. B**, 2018.

Anexo 01: Código utilizado no treinamento da rede neural

```
# Importação de bibliotecas

import pandas as pd
from tensorflow.keras import layers, models, optimizers
from tensorflow.keras.applications import VGG16
from tensorflow.keras.callbacks import EarlyStopping, ReduceLROnPlateau
from tensorflow.keras.metrics import FalseNegatives, FalsePositives, TrueN
egatives, TruePositives
from tensorflow.keras.preprocessing.image import ImageDataGenerator

# Construção da rede neural com base convolucional da rede VGG-16 e classi
ficador próprio

conv_base = VGG16(weights='imagenet', include_top=False, input_shape=(256,
256,3))
conv_base.trainable = False

model = models.Sequential()
model.add(conv_base)
model.add(layers.Flatten())
model.add(layers.Dense(512, activation='relu'))
model.add(layers.Dropout(0.5))
model.add(layers.Dense(1, activation='sigmoid'))

# Compilação da rede neural com especificação das métricas de avaliação

model.compile(
    loss='binary_crossentropy',
    optimizer=optimizers.Adam(),
    metrics=[FalseNegatives(), FalsePositives(), TrueNegatives(), TruePosi
tives()]
)

# Definição das funções de retorno

early_stop = EarlyStopping(
    monitor='val_true_negatives',
    min_delta=100,
    patience=5,
    restore_best_weights=True,
    mode='max',
    verbose=1
)

reduce_lr = ReduceLROnPlateau(
    monitor='val_loss',
    factor=0.2,
    patience=3,
    verbose=1,
    mode='min',
    min_delta=0.0001,
```

```
    min_lr=0
)

# Leitura dos arquivo CSV com as listas de imagens dos grupos de treinamen
to, validação e teste

df_train = df = pd.read_csv('train_data.csv')
df_val = df = pd.read_csv('validation_data.csv')
df_test = df = pd.read_csv('test_data.csv')

x_col = 'Image Index'
y_col = 'No Finding'

# Criação do conjunto de treinamento

train_datagen = ImageDataGenerator(rescale=1./255)

train_generator = train_datagen.flow_from_dataframe(
    dataframe=df_train,
    directory='./train',
    x_col=x_col,
    y_col=y_col,
    target_size=(256,256),
    color_mode='rgb',
    class_mode='raw',
    batch_size=16
)

# Criação do conjunto de validação

test_datagen = ImageDataGenerator(rescale=1./255)

validation_generator = test_datagen.flow_from_dataframe(
    dataframe=df_val,
    directory='./validation',
    x_col=x_col,
    y_col=y_col,
    target_size=(256,256),
    color_mode='rgb',
    class_mode='raw',
    batch_size=16
)

# Criação do conjunto de teste

test_generator = test_datagen.flow_from_dataframe(
    dataframe=df_test,
    directory='./test',
    x_col=x_col,
    y_col=y_col,
    target_size=(256,256),
    color_mode='rgb',
    class_mode='raw',
    batch_size=16
)
```

```
# Primeira etapa de treinamento

history = model.fit(
    train_generator,
    steps_per_epoch=60928/16,
    epochs=20,
    validation_data=validation_generator,
    validation_steps=25596/16,
    callbacks=[early_stop, reduce_lr]
)

Epoch 1/20
3808/3808 [==============================] - 1163s 303ms/step - loss: 0.68
55 - false_negatives: 3060.0000 - false_positives: 22160.0000 - true_negat
ives: 3255.0000 - true_positives: 32453.0000 - val_loss: 0.6381 - val_fals
e_negatives: 64.0000 - val_false_positives: 10470.0000 - val_true_negative
s: 139.0000 - val_true_positives: 14923.0000
Epoch 2/20
3808/3808 [==============================] - 1140s 299ms/step - loss: 0.64
88 - false_negatives: 5964.0000 - false_positives: 18871.0000 - true_negat
ives: 6544.0000 - true_positives: 29549.0000 - val_loss: 0.6270 - val_fals
e_negatives: 3176.0000 - val_false_positives: 5153.0000 - val_true_negativ
es: 5456.0000 - val_true_positives: 11811.0000
Epoch 3/20
3808/3808 [==============================] - 1143s 300ms/step - loss: 0.64
54 - false_negatives: 15388.0000 - false_positives: 8652.0000 - true_negat
ives: 16763.0000 - true_positives: 20125.0000 - val_loss: 0.6246 - val_fal
se_negatives: 2969.0000 - val_false_positives: 5482.0000 - val_true_negati
ves: 5127.0000 - val_true_positives: 12018.0000
Epoch 4/20
3808/3808 [==============================] - 1149s 302ms/step - loss: 0.64
26 - false_negatives: 15155.0000 - false_positives: 8678.0000 - true_negat
ives: 16737.0000 - true_positives: 20358.0000 - val_loss: 0.6269 - val_fal
se_negatives: 4783.0000 - val_false_positives: 3686.0000 - val_true_negati
ves: 6923.0000 - val_true_positives: 10204.0000
Epoch 5/20
3808/3808 [==============================] - 1152s 303ms/step - loss: 0.64
11 - false_negatives: 15794.0000 - false_positives: 7859.0000 - true_negat
ives: 17556.0000 - true_positives: 19719.0000 - val_loss: 0.6226 - val_fal
se_negatives: 3854.0000 - val_false_positives: 4433.0000 - val_true_negati
ves: 6176.0000 - val_true_positives: 11133.0000
Epoch 6/20
3808/3808 [==============================] - 1154s 303ms/step - loss: 0.64
01 - false_negatives: 16071.0000 - false_positives: 7580.0000 - true_negat
ives: 17835.0000 - true_positives: 19442.0000 - val_loss: 0.6196 - val_fal
se_negatives: 3641.0000 - val_false_positives: 4612.0000 - val_true_negati
ves: 5997.0000 - val_true_positives: 11346.0000
Epoch 7/20
3808/3808 [==============================] - 1154s 303ms/step - loss: 0.63
62 - false_negatives: 16134.0000 - false_positives: 7284.0000 - true_negat
ives: 18131.0000 - true_positives: 19379.0000 - val_loss: 0.6325 - val_fal
se_negatives: 5285.0000 - val_false_positives: 3319.0000 - val_true_negati
ves: 7290.0000 - val_true_positives: 9702.0000
Epoch 8/20
```

```
3808/3808 [==============================] - 1154s 303ms/step - loss: 0.63
52 - false_negatives: 16377.0000 - false_positives: 7078.0000 - true_negat
ives: 18337.0000 - true_positives: 19136.0000 - val_loss: 0.6171 - val_fal
se_negatives: 3663.0000 - val_false_positives: 4568.0000 - val_true_negati
ves: 6041.0000 - val_true_positives: 11324.0000
Epoch 9/20
3808/3808 [==============================] - 1155s 303ms/step - loss: 0.63
56 - false_negatives: 16669.0000 - false_positives: 6883.0000 - true_negat
ives: 18532.0000 - true_positives: 18844.0000 - val_loss: 0.6168 - val_fal
se_negatives: 3587.0000 - val_false_positives: 4613.0000 - val_true_negati
ves: 5996.0000 - val_true_positives: 11400.0000
Epoch 10/20
3808/3808 [==============================] - 1153s 303ms/step - loss: 0.63
26 - false_negatives: 16381.0000 - false_positives: 6924.0000 - true_negat
ives: 18491.0000 - true_positives: 19132.0000 - val_loss: 0.6279 - val_fal
se_negatives: 5606.0000 - val_false_positives: 3045.0000 - val_true_negati
ves: 7564.0000 - val_true_positives: 9381.0000
Epoch 11/20
3808/3808 [==============================] - 1150s 302ms/step - loss: 0.63
07 - false_negatives: 16591.0000 - false_positives: 6690.0000 - true_negat
ives: 18725.0000 - true_positives: 18922.0000 - val_loss: 0.6164 - val_fal
se_negatives: 3825.0000 - val_false_positives: 4412.0000 - val_true_negati
ves: 6197.0000 - val_true_positives: 11162.0000
Epoch 12/20
3808/3808 [==============================] - 1148s 301ms/step - loss: 0.62
92 - false_negatives: 16506.0000 - false_positives: 6602.0000 - true_negat
ives: 18813.0000 - true_positives: 19007.0000 - val_loss: 0.6171 - val_fal
se_negatives: 3745.0000 - val_false_positives: 4494.0000 - val_true_negati
ves: 6115.0000 - val_true_positives: 11242.0000
Epoch 13/20
3808/3808 [==============================] - 1152s 303ms/step - loss: 0.62
84 - false_negatives: 16635.0000 - false_positives: 6487.0000 - true_negat
ives: 18928.0000 - true_positives: 18878.0000 - val_loss: 0.6161 - val_fal
se_negatives: 2596.0000 - val_false_positives: 5743.0000 - val_true_negati
ves: 4866.0000 - val_true_positives: 12391.0000
Epoch 14/20
3808/3808 [==============================] - 1148s 301ms/step - loss: 0.62
53 - false_negatives: 17086.0000 - false_positives: 6081.0000 - true_negat
ives: 19334.0000 - true_positives: 18427.0000 - val_loss: 0.6314 - val_fal
se_negatives: 5449.0000 - val_false_positives: 3138.0000 - val_true_negati
ves: 7471.0000 - val_true_positives: 9538.0000
Epoch 15/20
3808/3808 [==============================] - 1149s 302ms/step - loss: 0.62
42 - false_negatives: 16777.0000 - false_positives: 6118.0000 - true_negat
ives: 19297.0000 - true_positives: 18736.0000 - val_loss: 0.6161 - val_fal
se_negatives: 3808.0000 - val_false_positives: 4420.0000 - val_true_negati
ves: 6189.0000 - val_true_positives: 11179.0000
Restoring model weights from the end of the best epoch.
Epoch 00015: early stopping
```

```
# Avaliação do primeiro treinamento no conjunto de teste

test_result = model.evaluate(test_generator, steps=25596/16)
```

```
1599/1599 [==============================] - 338s 211ms/step - loss: 0.675
0 - false_negatives: 6793.0000 - false_positives: 1725.0000 - true_negativ
es: 14010.0000 - true_positives: 3068.0000

# Habilitação da última camada convolucional para treinamento

conv_base.trainable = True

set_trainable = False
for layer in conv_base.layers:
    if layer.name == 'block5_conv3':
        set_trainable = True
    if set_trainable:
        layer.trainable = True
    else:
        layer.trainable = False

# Recompilação da rede neural

model.compile(
    loss='binary_crossentropy',
    optimizer=optimizers.Adam(learning_rate=1e-5),
    metrics=[FalseNegatives(), FalsePositives(), TrueNegatives(), TruePosi
tives()]
)

# Segunda etapa de treinamento

history = model.fit(
    train_generator,
    steps_per_epoch=60928/16,
    epochs=20,
    validation_data=validation_generator,
    validation_steps=25596/16,
    callbacks=[early_stop, reduce_lr]
)

Epoch 1/20
3808/3808 [==============================] - 1160s 300ms/step - loss: 0.62
54 - false_negatives: 15683.0000 - false_positives: 6944.0000 - true_negat
ives: 18471.0000 - true_positives: 19830.0000 - val_loss: 0.6178 - val_fal
se_negatives: 2999.0000 - val_false_positives: 5522.0000 - val_true_negati
ves: 5087.0000 - val_true_positives: 11988.0000
Epoch 2/20
3808/3808 [==============================] - 1141s 300ms/step - loss: 0.61
96 - false_negatives: 12386.0000 - false_positives: 9322.0000 - true_negat
ives: 16093.0000 - true_positives: 23127.0000 - val_loss: 0.6180 - val_fal
se_negatives: 3434.0000 - val_false_positives: 4818.0000 - val_true_negati
ves: 5791.0000 - val_true_positives: 11553.0000
Epoch 3/20
3808/3808 [==============================] - 1145s 301ms/step - loss: 0.61
07 - false_negatives: 10605.0000 - false_positives: 10071.0000 - true_nega
tives: 15344.0000 - true_positives: 24908.0000 - val_loss: 0.6121 - val_fa
lse_negatives: 2764.0000 - val_false_positives: 5514.0000 - val_true_negat
ives: 5095.0000 - val_true_positives: 12223.0000
```

```
Epoch 4/20
3808/3808 [==============================] - 1146s 301ms/step - loss: 0.59
91 - false_negatives: 10187.0000 - false_positives: 9848.0000 - true_negat
ives: 15567.0000 - true_positives: 25326.0000 - val_loss: 0.6076 - val_fal
se_negatives: 3849.0000 - val_false_positives: 4491.0000 - val_true_negati
ves: 6118.0000 - val_true_positives: 11138.0000
Epoch 5/20
3808/3808 [==============================] - 1144s 300ms/step - loss: 0.58
56 - false_negatives: 9723.0000 - false_positives: 9589.0000 - true_negati
ves: 15826.0000 - true_positives: 25790.0000 - val_loss: 0.6057 - val_fals
e_negatives: 4303.0000 - val_false_positives: 3985.0000 - val_true_negativ
es: 6624.0000 - val_true_positives: 10684.0000
Epoch 6/20
3808/3808 [==============================] - 1175s 309ms/step - loss: 0.57
80 - false_negatives: 9183.0000 - false_positives: 9764.0000 - true_negati
ves: 15651.0000 - true_positives: 26330.0000 - val_loss: 0.6009 - val_fals
e_negatives: 3672.0000 - val_false_positives: 4387.0000 - val_true_negativ
es: 6222.0000 - val_true_positives: 11315.0000
Epoch 7/20
3808/3808 [==============================] - 1163s 305ms/step - loss: 0.56
78 - false_negatives: 8924.0000 - false_positives: 9520.0000 - true_negati
ves: 15895.0000 - true_positives: 26589.0000 - val_loss: 0.6018 - val_fals
e_negatives: 3745.0000 - val_false_positives: 4346.0000 - val_true_negativ
es: 6263.0000 - val_true_positives: 11242.0000
Epoch 8/20
3808/3808 [==============================] - 1147s 301ms/step - loss: 0.56
17 - false_negatives: 8790.0000 - false_positives: 9484.0000 - true_negati
ves: 15931.0000 - true_positives: 26723.0000 - val_loss: 0.6049 - val_fals
e_negatives: 3081.0000 - val_false_positives: 4964.0000 - val_true_negativ
es: 5645.0000 - val_true_positives: 11906.0000
Epoch 9/20
3808/3808 [==============================] - 1144s 300ms/step - loss: 0.55
28 - false_negatives: 8595.0000 - false_positives: 9268.0000 - true_negati
ves: 16147.0000 - true_positives: 26918.0000 - val_loss: 0.6048 - val_fals
e_negatives: 4155.0000 - val_false_positives: 3993.0000 - val_true_negativ
es: 6616.0000 - val_true_positives: 10832.0000

Epoch 00009: ReduceLROnPlateau reducing learning rate to 1.999999949475750
5e-06.
Epoch 10/20
3808/3808 [==============================] - 1157s 304ms/step - loss: 0.53
47 - false_negatives: 8477.0000 - false_positives: 8894.0000 - true_negati
ves: 16521.0000 - true_positives: 27036.0000 - val_loss: 0.6104 - val_fals
e_negatives: 3121.0000 - val_false_positives: 4804.0000 - val_true_negativ
es: 5805.0000 - val_true_positives: 11866.0000
Restoring model weights from the end of the best epoch.
Epoch 00010: early stopping
```

```
# Avaliação do segundo treinamento no conjunto de teste

test_result = model.evaluate(test_generator, steps=25596/16)
```

```
1599/1599 [==============================] - 340s 212ms/step - loss:
0.6249 - false_negatives: 5758.0000 - false_positives: 2470.0000 -
true_negatives: 13265.0000 - true_positives: 4103.0000
```