

# Uma aplicação para reconhecimento de carros utilizando Deep Learning

Felipe Nakandakari dos Santos<sup>1</sup>, Luigi Uematsu<sup>2</sup>

Ivan Carlos Alcântara de Oliveira

<sup>1</sup>Curso de Ciência da Computação

Universidade Presbiteriana Mackenzie (UPM) – São Paulo, SP - Brasil

<sup>2</sup>Curso de Ciência da Computação

Universidade Presbiteriana Mackenzie (UPM) – São Paulo, SP - Brasil

[10395160@mackenzista.com.br](mailto:10395160@mackenzista.com.br), [10396148@mackenzista.com.br](mailto:10396148@mackenzista.com.br)

**Abstract.** *This work proposes the development of an application that recognizes vehicles through images captured by a mobile device camera and provides their technical specifications in real time. The implementation of this solution will require the application of advanced Artificial Intelligence techniques, such as machine learning and deep learning, for the development of robust and accurate image recognition algorithms. The expectation is that the application will contribute to the creation of a useful and innovative tool for automotive enthusiasts, professionals in the area, and consumers in general.*

**Keywords:** *image recognition, artificial intelligence, deep learning, mobile application, vehicle identification.*

**Resumo.** *Este trabalho propõe o desenvolvimento de um aplicativo que reconhece veículos por meio de imagens capturadas por uma câmera de um dispositivo móvel e fornece a sua ficha técnica em tempo real. A implementação dessa solução exigirá a aplicação de técnicas avançadas de Inteligência Artificial, como aprendizado de máquina e deep learning, para o desenvolvimento de algoritmos robustos e precisos de reconhecimento de imagens. A expectativa é que o aplicativo contribua para a criação de uma ferramenta útil e inovadora para entusiastas de automóveis, profissionais da área e consumidores em geral.*

**Palavras-chave:** *reconhecimento de imagem, inteligência artificial, deep learning, aplicativo móvel, identificação de veículos.*

## **1. Introdução**

A convergência entre tecnologia e cotidiano tem remodelado nossa percepção e interação com o mundo. A ubiquidade de dispositivos móveis e a sofisticação crescente da inteligência artificial impulsionam a criação de soluções inovadoras em diversos setores. Conforme destacado por LeCun, Bengio e Hinton (2015), a capacidade das máquinas de compreender e interpretar o mundo visual, por meio do reconhecimento de imagens, tem se mostrado fundamental para a criação de aplicações mais intuitivas e personalizadas.

### **1.1 Contextualização**

A crescente popularidade de dispositivos móveis e o avanço das tecnologias de inteligência artificial (IA) têm impulsionado a criação de aplicativos inovadores que facilitam a vida das pessoas e transformam a forma como interagimos com o mundo ao nosso redor. Uma área que tem se beneficiado significativamente dessa sinergia é a de reconhecimento de imagens, que permite a computadores interpretar e compreender o conteúdo visual de fotos e vídeos.

O reconhecimento de veículos, em particular, representa um campo de pesquisa com grande potencial de aplicação em diversas áreas. Desde a segurança pública, com o auxílio na identificação de veículos roubados ou envolvidos em crimes, até o setor automotivo, com a criação de sistemas de auxílio ao motorista mais inteligentes, a capacidade de identificar um veículo a partir de uma imagem pode trazer inúmeros benefícios.

Neste contexto, este trabalho propõe o desenvolvimento de um protótipo de aplicação que tem a capacidade de reconhecer veículos por meio de imagens capturadas por uma câmera de um dispositivo móvel, fornecendo informações detalhadas sobre eles em tempo real. A ideia é que o usuário possa apontar a câmera do seu dispositivo para um carro e, em tempo real, obter informações como marca, modelo, ano de fabricação e características técnicas.

## 1.2 Justificativa

A crescente popularização de smartphones e o avanço da inteligência artificial (IA) propiciam o desenvolvimento de soluções inovadoras que transformam a maneira como interagimos com o mundo ao nosso redor. Neste contexto, o desenvolvimento de um aplicativo móvel capaz de identificar veículos a partir de imagens surge como uma proposta promissora, impulsionada por diversos fatores.

Em primeiro lugar, a democratização do acesso à informação é um dos principais motivadores deste projeto. Ao colocar nas mãos dos usuários uma ferramenta que permite identificar qualquer veículo, contribui-se para a disseminação de conhecimento sobre o universo automotivo. Essa democratização do acesso à informação possibilita que entusiastas, profissionais e consumidores em geral realizem pesquisas mais aprofundadas sobre modelos específicos, auxiliando na tomada de decisões mais assertivas, como a compra de um novo veículo.

Além disso, a implementação de um aplicativo de identificação de veículos exige a aplicação de técnicas avançadas de IA, como o aprendizado de máquina (*machine learning*), aprendizado profundo (*deep learning*) e o processamento de imagens. Desse modo, o desenvolvimento deste projeto contribui para o avanço da pesquisa nesta área e para a criação de novas soluções inovadoras que podem ser aplicadas em diversos outros contextos.

Por fim, o potencial de mercado para este tipo de aplicativo é considerável. A crescente demanda por soluções tecnológicas que facilitem a vida das pessoas indica um grande interesse por ferramentas que ofereçam informações precisas e detalhadas sobre veículos de forma rápida e intuitiva.

Em suma, o desenvolvimento de um aplicativo de identificação de veículos por imagem representa uma oportunidade única de combinar tecnologia e conhecimento para criar uma solução útil e inovadora, que atende a uma demanda crescente do mercado e contribui para o avanço da pesquisa em IA.

### 1.3 Objetivo Geral

Cada vez mais pessoas têm se interessado pelo meio automobilístico, querendo se aprofundar mais, porém uma das dificuldades para aqueles que estão entrando nessa área, é reconhecer os diversos modelos de carros existentes. Pensando nisso surgiu a ideia de desenvolver uma aplicação com a finalidade de ajudar o reconhecimento do modelo de um carro e suas características. A ideia central é que uma pessoa com o seu celular possa tirar uma foto e o sistema reconheça o modelo presente na imagem retornando suas especificações, por exemplo: nome da marca, modelo e ano de fabricação.

#### 1.3.1 Objetivos Específicos

Em relação aos objetivos específicos do trabalho, têm-se:

- Revisão de trabalhos semelhantes
- Análise e tratamento do *dataset base*, realizando a limpeza, treino e teste das imagens.
- Estudo de técnicas para treino e teste e implementação da melhor técnica
- Captura de imagens de veículos em tempo real com a utilização da câmera do dispositivo móvel para capturar imagens de veículos.
- Identificação do veículo por meio de algoritmos de reconhecimento de imagem para identificar a marca, modelo, ano de fabricação e outras características do veículo.
- Ser intuitivo e fácil de usar: Possuir uma interface amigável que permita que qualquer usuário possa utilizar o aplicativo sem dificuldades.

## 2. Referencial Teórico

Este capítulo se destina a uma abordagem mais aprofundada dos temas pertinentes ao trabalho, incluindo conceitos importantes, tais como visão computacional e *deep learning*. Além disso, serão discutidas as características e o funcionamento do modelo de rede neural convolucional (CNN) empregado para a identificação do modelo conforme a imagem. Para embasar academicamente o trabalho, serão referenciados estudos

relevantes e dados estatísticos relacionados ao tema proposto, contribuindo para uma fundamentação sólida e consistente.

## **2.1 Visão Computacional**

A Visão Computacional, um ramo da inteligência artificial, confere aos computadores a habilidade de "ver" e interpretar o mundo visual. Através de algoritmos e modelos matemáticos, as máquinas processam imagens e vídeos, extraindo informações relevantes e tomando decisões com base nessas informações. Essa área tem experimentado um crescimento exponencial, impulsionada por avanços em hardware, algoritmos e pela disponibilidade de grandes volumes de dados.

A Visão Computacional busca mimetizar a complexa capacidade humana de perceber o mundo visual. Para isso, ela se baseia em diversos conceitos e técnicas, como aquisição de imagens, pré-processamento, extração de características, reconhecimento de padrões e interpretação e análise.

As aplicações da Visão Computacional são vastas e abrangem diversas áreas, como reconhecimento facial, detecção de objetos, segmentação de imagens, análise de vídeo, realidade aumentada, medicina e indústria.

O surgimento do *deep learning*, um subcampo do aprendizado de máquina, revolucionou a Visão Computacional. As redes neurais convolucionais (CNNs) se tornaram a ferramenta de escolha para diversas tarefas, como classificação de imagens, detecção de objetos e segmentação. As CNNs são capazes de aprender representações hierárquicas de dados visuais, extraindo características de baixo nível até características de alto nível.

Apesar dos grandes avanços, a Visão Computacional ainda enfrenta desafios, como a variação nas condições de iluminação, a oclusão de objetos, a grande quantidade de dados necessária para treinar modelos e a interpretabilidade dos resultados. As

tendências futuras incluem o aprendizado por reforço, a visão computacional 3D e a visão computacional explicável.

A Visão Computacional é um campo em constante evolução, com um potencial enorme para transformar diversas áreas da nossa sociedade. A combinação de algoritmos sofisticados, hardware poderoso e grandes volumes de dados está impulsionando o desenvolvimento de aplicações cada vez mais inovadoras e eficientes.

## **2.2 Deep Learning**

O *deep learning*, ou aprendizado profundo, é um subcampo do aprendizado de máquina que utiliza redes neurais artificiais com múltiplas camadas para aprender representações complexas de dados. Inspirado no funcionamento do cérebro humano, o *deep learning* tem revolucionado a forma como abordamos problemas em diversas áreas, como visão computacional, processamento de linguagem natural e reconhecimento de fala.

As redes neurais artificiais, a base do *deep learning*, são compostas por camadas de neurônios artificiais interconectados. Cada neurônio realiza uma operação matemática simples, e a combinação dessas operações em várias camadas permite que a rede aprenda padrões complexos nos dados. O processo de aprendizado envolve a apresentação de grandes quantidades de dados à rede, ajustando os pesos das conexões entre os neurônios de forma a minimizar o erro entre a saída prevista pela rede e a saída real.

As aplicações do *deep learning* são vastas e abrangem diversas áreas, como reconhecimento de imagens, detecção de objetos, segmentação de imagens, geração de imagens, tradução automática, análise de sentimentos, geração de texto, *chatbots*, transcrição de áudio, comandos de voz, sistemas de diálogo, análise de sequências genéticas, descoberta de drogas, previsão de séries temporais, detecção de fraudes e desenvolvimento de agentes inteligentes para jogos.

Existem diversos tipos de redes neurais utilizadas em *deep learning*, cada uma com suas características e aplicações específicas, como redes neurais convolucionais

(CNNs), redes neurais recorrentes (RNNs), redes neurais generativas adversariais (GANs) e transformers.

Apesar de seus sucessos, o *deep learning* ainda enfrenta desafios, como a necessidade de grandes quantidades de dados para treinamento, a interpretabilidade dos modelos e a questão da privacidade dos dados. As tendências futuras do deep learning incluem *deep learning* federado, explicabilidade dos modelos e *deep learning* em dispositivos móveis.

O *deep learning* é uma área em rápida evolução, com um potencial enorme para transformar diversas áreas da nossa sociedade. A combinação de algoritmos sofisticados, hardware poderoso e grandes volumes de dados está impulsionando o desenvolvimento de aplicações cada vez mais inovadoras e eficientes.

## **2.3 Redes Neurais Convolucionais**

As redes neurais convolucionais (CNNs, do inglês Convolutional Neural Networks) são um tipo específico de rede neural artificial que se destaca no processamento de dados visuais. Elas são amplamente utilizadas em tarefas de visão computacional, como reconhecimento de imagens, detecção de objetos, segmentação e geração de imagens.

Inspiradas no funcionamento do córtex visual humano, as CNNs extraem características hierárquicas de uma imagem, começando por características simples como bordas e texturas e progredindo para características mais complexas como objetos inteiros. Essa capacidade é alcançada através de camadas convolucionais, que aplicam filtros (kernels) sobre a imagem de entrada para extrair características locais, e camadas de *pooling*, que reduzem a dimensionalidade da representação. As saídas dessas camadas são então processadas por camadas totalmente conectadas, que são responsáveis pela classificação ou regressão final.

As CNNs são particularmente eficientes para imagens devido à sua invariância a translações, extração de características hierárquicas e compartilhamento de pesos. Essas características as tornam ideais para uma ampla gama de aplicações, como reconhecimento de imagens, detecção de objetos, segmentação de imagens, geração de imagens, análise de vídeo e muitas outras.

Apesar de seu sucesso, as CNNs ainda enfrentam desafios, como a necessidade de grandes quantidades de dados para treinamento, a interpretabilidade dos modelos e a computação intensiva. As tendências futuras incluem o desenvolvimento de arquiteturas mais eficientes, a aplicação de CNNs para dados 3D e a utilização de CNNs para tarefas mais complexas, como compreensão de cenas e raciocínio visual.

Em resumo, as redes neurais convolucionais são uma ferramenta poderosa para o processamento de imagens e têm revolucionado a área de visão computacional. Sua capacidade de extrair características hierárquicas e invariantes a translações as torna ideais para uma ampla gama de aplicações.

As redes neurais convolucionais (CNNs) são uma ferramenta poderosa para o processamento de dados visuais, e sua aplicação na classificação de imagens de carros é um exemplo clássico. Ao serem treinadas com uma base de dados diversificada de imagens de carros, as CNNs aprendem a extrair características relevantes, como forma, cor, textura e componentes, para distinguir entre diferentes marcas, modelos e anos.

## **2.4 Trabalhos Relacionados**

Nesta subseção serão apresentados trabalhos/projetos relacionados com a proposta deste trabalho. O projeto "*Identify-a-Car-Model-with-Deep-Learning*", disponível em <https://github.com/ajayrawatsap/Identify-a-Car-Model-with-Deep-Learning>, oferece um guia prático para aplicar técnicas de aprendizado profundo na identificação de modelos de carros. O autor demonstra como coletar, organizar e processar dados de imagens de carros para treinar modelos de *deep learning* capazes de classificar diferentes



marcas e modelos de veículos. O projeto utiliza as bibliotecas *Keras* e *PyTorch* em Python, explorando técnicas avançadas como aumento de dados, *dropout* e aprendizado de transferência.

Rawat et al. (2021), propõe um sistema eficiente para detecção e reconhecimento de veículos utilizando técnicas avançadas de aprendizado profundo, com aplicação em visão computacional. O modelo combina o algoritmo *YOLOv3*, conhecido por sua velocidade e precisão na detecção de objetos, com a arquitetura *MobileNet* para classificação dos veículos detectados em categorias específicas, como grandes (SUVs e minivans) e pequenos (sedans e *hatchbacks*). Treinado com os conjuntos de dados COCO e *Stanford Car Dataset*, além de dados suplementares obtidos por *web scraping*, o modelo alcançou resultados superiores em métricas como precisão média (mAP) e taxa de quadros por segundo (FPS), quando comparado a modelos de referência como SSD e *Faster R-CNN*, apresentando um aumento de 15% na precisão média. Além de demonstrar robustez e generalização em cenários diversos, o sistema pode ser expandido para funcionalidades como estimativa de velocidade e prevenção de colisões, destacando-se como uma contribuição relevante para sistemas avançados de assistência ao motorista (ADAS) e outras aplicações em visão computacional.

### **3. Metodologia de pesquisa**

Para apresentar uma visão abrangente e estruturada do progresso deste trabalho, esta seção descreve as atividades concluídas, os materiais e métodos implementados até agora. O desenvolvimento do projeto é construído usando métodos de *deep learning* aplicado ao problema do reconhecimento de modelos de carro. O *dataset* escolhido neste trabalho contém fotos que mostram diferentes modelos de carro, suas cores e anos de produção, podendo ser encontrado em: <https://deepvisualmarketing.github.io>.

### 3.1 Seleção do *dataset*:

Foi escolhido um *dataset* que categoriza veículos através de marca, modelo, cor e ano de fabricação, com fotos tiradas de diferentes ângulos. Considerando a importância de ter um *dataset* que abrange uma vasta quantidade de veículos, desde modelos de carros populares até aqueles mais raros para garantir que o modelo seja capaz de generalizar para diferentes tipos de veículos, o *dataset* escolhido possui um total de 1451784 imagens de 899 modelos de carros. Para os primeiros testes, escolhemos dois modelos iniciais: Honda Civic, com 2.041 imagens, e Audi Q3, com 4.817 imagens, considerando as cores preto, cinza e branco.

### 3.2 Desenvolvimento do modelo inicial:

Um código em Python foi desenvolvido com as bibliotecas *TensorFlow*, *Keras* e *ImageDataGenerator*. O modelo se baseia em uma CNN sequencial por camadas de *pooling*, uma camada de *flattening* e camadas densas. O algoritmo utilizado para o treino é um modelo base da CNN no *Keras TensorFlow*, com a função de ativação ‘relu’ para camada convolucional densa, e *softmax* para camada de saída e geração de probabilidades para cada classe (cor). Com isto, foi utilizado o *dataset* ‘DVM-CAR’ feito por Jingmin Huang e Bowi Chen para realizar experimentos com a rede neural convolucional básica do *tensorflow keras*.

Posteriormente pretende-se utilizar um modelo de CNN pré-treinada, como por exemplo: *ResNet-50*, *VGG* e *Vision Transformers* para treinarmos nossa IA. Com o melhor modelo obtido, será iniciado o desenvolvimento de uma aplicação que seja capaz de reconhecer modelos de carros e assim poder realizar testes para validar a eficácia do modelo.

## 4. Resultados

Foram escolhidos dois modelos do *dataset* 'DVM-CAR': Audi Q3 e Honda Civic. O treinamento e teste dos modelos Civic e Q3 foram realizados na plataforma Google Colab, utilizando a biblioteca *Keras* integrada ao framework *TensorFlow* em linguagem *Python*. O Colab foi escolhido para aproveitar recursos de processamento, como GPUs gratuitas, otimizando o desempenho durante o treinamento. As imagens utilizadas foram organizadas em pastas estruturadas por ano e cor dos veículos, com a criação de um *DataFrame* para associar os caminhos das imagens às respectivas classes. Para melhorar a generalização dos modelos e evitar *overfitting*, foi utilizado o *ImageDataGenerator*, que aplicou *Data Augmentation* com transformações como rotações, zoom, espelhamento e ajustes de escala.

Ambos os modelos foram implementados com uma arquitetura de rede neural convolucional (CNN) sequencial. As redes incluíram camadas convolucionais para extração de características, camadas de *pooling* para redução de dimensionalidade, uma camada de *flattening* para transformar os mapas de características em vetores unidimensionais e camadas densas para a classificação final. A função de ativação '*relu*' foi utilizada nas camadas ocultas, e a função *Softmax* foi aplicada na camada de saída para gerar probabilidades das classes (cores dos veículos). A compilação foi realizada com a função de perda *categorical\_crossentropy*, adequada para classificação multiclasse, e o otimizador *Adam*, conhecido por sua eficiência no ajuste dos pesos. A métrica de avaliação escolhida foi a acurácia.

Os treinamentos foram realizados em 10 épocas para os dois modelos, com monitoramento contínuo das métricas de perda e acurácia nos conjuntos de treinamento e validação. O modelo Civic mostrou uma tendência consistente de aumento na acurácia e redução na perda, indicando um aprendizado eficiente e boa generalização. O modelo Q3 apresentou um desempenho semelhante, com algumas diferenças no comportamento das métricas, o que sugere que ajustes específicos podem ser necessários para diferentes arquiteturas ou tipos de dados.

Os gráficos gerados para ambos os modelos indicaram que a utilização de técnicas de *Data Augmentation*, aliada a uma arquitetura de CNN otimizada, resultou em um bom equilíbrio entre aprendizado e generalização. Testes comparativos demonstraram que ambos os modelos são robustos e capazes de classificar novas imagens de veículos com precisão, sendo o desempenho avaliado de forma individual para identificar pontos fortes e oportunidades de melhoria em cada abordagem.

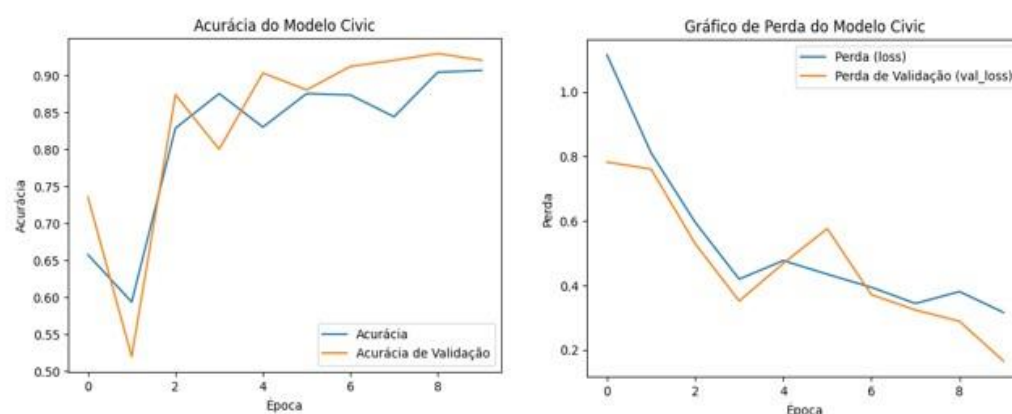


Figura 1: Gráficos de Acurácia e Loss ao longo de 10 épocas para o Civic.

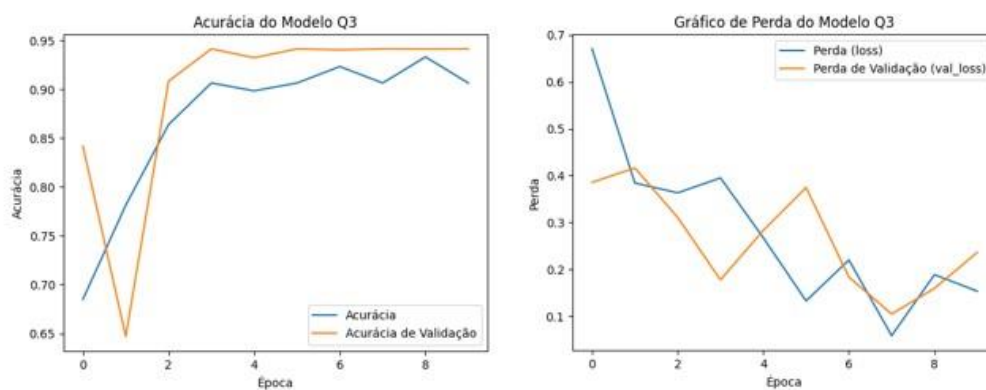


Figura 2: Gráficos de Acurácia e Loss ao longo de 10 épocas para o Q3.

O gráfico de acurácia apresenta a evolução da acurácia dos modelos ao longo das épocas durante o treinamento e validação. A linha azul representa a acurácia no conjunto de treinamento, enquanto a linha laranja indica a acurácia no conjunto de validação. Observa-se que ambas as curvas mostram uma tendência de crescimento, o que indica que o modelo está aprendendo de forma consistente. A acurácia no conjunto de validação é ligeiramente superior à do treinamento, sugerindo uma boa capacidade de generalização do modelo e ausência de *overfitting* significativo. Ao final do treinamento, ambas as curvas convergem para valores próximos de 0,90, indicando um desempenho robusto.

O gráfico de perda exhibe a diminuição da perda (*loss*) ao longo das épocas para os conjuntos de treinamento e validação. A linha azul representa a perda no conjunto de treinamento, enquanto a linha laranja mostra a perda no conjunto de validação. Nota-se uma redução contínua na perda, evidenciando que o modelo está melhorando sua capacidade de prever corretamente as classes das imagens. Apesar de uma oscilação pontual na perda de validação, especialmente nas épocas intermediárias, as curvas mostram um comportamento consistente e convergente, com valores baixos de perda ao final do treinamento, reforçando a eficácia do modelo.

Esses gráficos ilustram que o modelo alcançou um bom equilíbrio entre aprendizado e generalização, indicando que ele é capaz de classificar com precisão novas imagens de veículos de maneira consistente.

## 5. Conclusão

Concluimos que as etapas propostas no projeto não foram integralmente concluídas devido à limitação de tempo disponível para sua execução. No entanto, os experimentos realizados apresentaram resultados satisfatórios, contribuindo significativamente para o aprimoramento da etapa de treinamento. Essa evolução não apenas facilita o processo de treino das imagens, mas também estabelece uma base sólida para a futura implementação da aplicação. Assim, ao alcançar um algoritmo eficiente para o treinamento, a etapa de implementação será consideravelmente simplificada, otimizando todo o desenvolvimento do projeto.

## 6. Endereço GitHub e YouTube

GitHub: <https://github.com/Felipe072502/Uma-aplicacao-para-reconhecimento-de-carros-utilizando-Deep-Learning>

YouTube: <https://youtu.be/6QJ24p2IW6g>

## 7. Referências

ARRIETA, Alejandro Barredo et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, v. 58, p. 82-115, 2020.

Acesso em: 08 set. 2024.

SAHOO, Kabita et al. Exploratory data analysis using Python. **International Journal of Innovative Technology and Exploring Engineering**, v. 8, n. 12, p. 4727-4735, 2019.

Acesso em: 08 set. 2024.

QIAN, Ling et al. Cloud computing: An overview. In: **Cloud Computing: First International Conference, CloudCom 2009, Beijing, China, December 1-4, 2009. Proceedings 1**. Springer Berlin Heidelberg, 2009. p. 626-631. Acesso em: 08 set. 2024.

**O que é o Deep Learning? | IBM**. Disponível em:

<<https://www.ibm.com/br-pt/topics/deep-learning>>. Acesso em: 08 set. 2024.

**Search|Kaggle**. Disponível em:

<<https://www.kaggle.com/search?q=car+images+dataset+in%3Adatasets>>. Acesso em: 12 set. 2024.

**CompCars Dataset**. Disponível em:

<[https://mmlab.ie.cuhk.edu.hk/datasets/comp\\_cars/](https://mmlab.ie.cuhk.edu.hk/datasets/comp_cars/)>. Acesso em: 12 set. 2024.

DVM Car Dataset: A Large-Scale Dataset for Automotive Applications. Disponível em:

<<https://deepvisualmarketing.github.io/>>. Acesso em: 15 out.2024

**Frota de Veículos - 2024.** Disponível em:

<<https://www.gov.br/transportes/pt-br/assuntos/transito/conteudo-Senatran/frota-de-veiculos-2024>>. Acesso em: 12 set.2024

**Sinesp Cidadão.** Disponível em:

<<https://www.gov.br/mj/pt-br/assuntos/sua-seguranca/seguranca-publica/sinesp-1/sinesp-Cidadao>>. Acesso em: 12 set. 2024.

YOSHUA BENGIO. Learning deep architectures for AI. Boston: Now, 2009.

SZELISKI, R. Computer vision : algorithms and applications. London: Springer, 2011.