

UNIDADE DE ENSINO SUPERIOR DOM BOSCO  
CURSO DE ENGENHARIA DE SOFTWARE

**CELSO LUIZ SILVA SOARES FILHO**

**CLASSIFICAÇÃO DE EMOÇÕES USANDO INTELIGÊNCIA ARTIFICIAL PARA  
MAIOR SEGURANÇA NO TRÂNSITO**

São Luís  
2023

Celso Luiz Silva Soares Filho

**CLASSIFICAÇÃO DE EMOÇÕES USANDO INTELIGÊNCIA ARTIFICIAL PARA  
MAIOR SEGURANÇA NO TRÂNSITO**

Monografia apresentada ao Curso de Graduação de Engenharia de Software da Unidade de Ensino Superior Dom Bosco como requisito parcial para a obtenção do grau de Bacharel em Engenharia de Software.

Orientador: Prof. Dr. Giovanni Lucca França da Silva

São Luís  
2023

Dados Internacionais de Catalogação na Publicação (CIP)  
Centro Universitário - UNDB / Biblioteca

Soares Filho, Celso Luiz Silva

Classificação de emoções usando inteligência artificial para maior segurança no trânsito. / Celso Luiz Silva Soares Filho. \_\_\_\_  
São Luís, 2023.

53 f.

Orientador: Prof. Dr. Giovanni Lucca França da Silva.  
Monografia (Graduação em Engenharia de Software) –  
Curso de Engenharia de Software - Centro Universitário  
Unidade de Ensino Superior Dom Bosco – UNDB, 2023.

1. Inteligência artificial. 2. Trânsito. 3. Sentimentos em vídeos. 4. Reconhecimento facial. 5. Segurança no trânsito.  
I. Título.

CDU 004.81:159.942.52

**CELSO LUIZ SILVA SOARES FILHO**

**CLASSIFICAÇÃO DE EMOÇÕES USANDO INTELIGÊNCIA ARTIFICIAL PARA  
MAIOR SEGURANÇA NO TRÂNSITO**

Monografia apresentada ao Curso de Engenharia de Software do Centro Universitário Unidade de Ensino Superior Dom Bosco como requisito parcial para obtenção do grau de Bacharel em Engenharia de Software.

Aprovada em: \_\_\_\_/\_\_\_\_/\_\_\_\_.

**BANCA EXAMINADORA:**

---

**Prof. Dr. Giovanni Lucca França da Silva**

(Orientador)

Doutor em Engenharia Elétrica

Centro universitário UNDB

---

**Prof. Me. Allison Jorge Silva Almeida**

Mestre em Inteligência Artificial

Centro Universitário Unidade de Ensino Superior Dom Bosco (UNDB)

---

**Prof. Me. Arlison Wandy Sousa Martins**

Mestre em Ciência da Computação

Centro Universitário Unidade de Ensino Superior Dom Bosco (UNDB)  
Bosco (UNDB)

*Dedico às pessoas que sempre  
estiveram meu lado nos momentos  
difíceis*

## **AGRADECIMENTOS**

A Deus pela sabedoria e principalmente a saúde. Ao meu primo Pedro, que me mostrou a área de tecnologia ainda quando eramos crianças. Ao meu primo Jhones que sempre acreditou no meu potencial. A minha namorada Maria Amanda, que esteve comigo durante todos os momentos difíceis e sempre acreditou e me deu força. Aos meus amigos Lucas, Luís e Pedro, pelas melhores risadas nos momentos em que eu mais precisava. Ao meu professor e orientador Giovanni, pela paciência e por ter me apresentado e conduzido durante todo meu processo de aprendizado com a área de inteligência artificial. A todos os meus professores. A todos os amigos da que conheci durante o curso, em especial, Carlos Laranjeira, Caique, Aranha, Gidalti e Juan. Ao meu amigo Enos, que sempre escutou e deu atenção quando eu falava sobre os assuntos de IA. E a todos que contribuíram para minha formação acadêmica.

## RESUMO

O desenvolvimento de carros inteligentes é uma tecnologia promissora para o século XXI, com a segurança sendo uma preocupação chave. Nesse contexto, a incorporação de técnicas avançadas de Inteligência Artificial e visão computacional, como o reconhecimento de emoções, surge como uma solução potencial para aprimorar a segurança no trânsito, permitindo a análise dos condutores e a prevenção de comportamentos perigosos. Portanto, o objetivo deste trabalho é utilizar uma rede neural convolucional (CNN) para a classificação de emoções em vídeos, visando a segurança no trânsito. A análise facial é usada para a extração das características relevantes para a classificação. Os resultados obtidos com a metodologia proposta demonstraram uma precisão média de 82,82%, sensibilidade média de 80,87% e acurácia média de 81,00% na classificação de emoções. Esses resultados reforçam a importância e o potencial das técnicas de reconhecimento de emoções para melhorar a segurança no trânsito e servir para futuros estudos sobre o tema.

**Palavras-chave:** Emoções. Classificação. Rede neural convolucional. Inteligência Artificial. Visão computacional.

## **ABSTRACT**

The development of smart cars is a promising technology for the 21st century, with safety being a key concern. In this context, the incorporation of advanced Artificial Intelligence and computer vision techniques, such as emotion recognition, emerges as a potential solution to enhance traffic safety by enabling driver analysis and preventing dangerous behaviors. Therefore, the objective of this work is to utilize a convolutional neural network (CNN) for emotion classification in videos, aiming at traffic safety. Facial analysis is used to extract relevant features for classification. The results obtained with the proposed methodology demonstrated an average precision of 82.82%, average sensitivity of 80.87%, and average accuracy of 81.00% in emotion classification. These results reinforce the importance and potential of emotion recognition techniques to improve traffic safety and contribute to future studies in the field.

**Keywords:** Emotions. Classification. Convolutional neural networks. Artificial Intelligence. Computer vision.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Áreas relacionadas com a IA .....	26
Figura 2 – Exemplo de rede neural de múltiplas camadas .....	28
Figura 3 – Rede convolucional -LeNet 5 .....	30
Figura 4 – Exemplo de aplicação do Kernel 3x3 .....	31
Figura 5 – Antes e depois da aplicação do dropout em uma rede neural .....	33
Figura 6 – Diagrama do método proposto .....	35
Figura 7 – Fluxo da arquitetura da rede neural .....	38
Figura 8 – Quantidade de imagens nos conjuntos de treino e validação .....	41
Figura 9 – Exemplo do processo de identificação de faces para classificação em novas imagens .....	42
Figura 10 – Matriz de confusão .....	43
Figura 11 – Matriz de confusão: observação da emoção raiva .....	43
Figura 12 – Classificação em diferentes etapas do sorriso .....	46
Figura 13 – Classificação diferente em imagens semelhantes .....	46
Figura 14 – Dificuldade na detecção do rosto por causa dos óculos .....	47
Figura 15 – Erros recorrentes nos testes feitos no período noturno com óculos .....	48
Figura 16 – Erros recorrentes nos testes feitos no período noturno sem óculos .....	48

## LISTA DE TABELAS

Tabela 1 - Métricas de avaliação .....	43
--	----

## **LISTA DE ABREVIATURAS E SIGLAS**

CNN - Rede neural convolucional

MIT – Instituto de Tecnologia de Massachusetts

IA - Inteligência Artificial

DL - Deep Learning

RNA - Rede neural artificial

ReLU – Rectified Linear Unit

FER-2013 – Facial Expression Recognition 2013

VP - Verdadeiro positivo

VN - Verdadeiro negativo

FN - Falso negativo

FP - Falso positivo

LBP – Local Binary Patterns

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>13</b>
<b>1.1</b>	<b>Justificativa .....</b>	<b>14</b>
<b>1.2</b>	<b>Objetivo geral .....</b>	<b>14</b>
<b>1.3</b>	<b>Organização do Trabalho .....</b>	<b>15</b>
<b>2</b>	<b>TRABALHOS RELACIONADOS .....</b>	<b>16</b>
<b>3</b>	<b>REFERENCIAL TEÓRICO .....</b>	<b>18</b>
<b>3.1</b>	<b>Detecção de faces .....</b>	<b>18</b>
<b>3.2</b>	<b>Identificação de sentimentos por expressões faciais .....</b>	<b>19</b>
<b>3.3</b>	<b>Tipos de emoções .....</b>	<b>21</b>
<b>3.4</b>	<b>Reconhecimento de sentimentos em carros Inteligentes .....</b>	<b>23</b>
<b>3.5</b>	<b>Imagens digitais .....</b>	<b>24</b>
<b>3.6</b>	<b>Inteligência Artificial .....</b>	<b>25</b>
3.6.1	Aprendizado de máquina profundo .....	27
3.6.1.1	Rede neural artificial .....	27
3.6.1.2	Redes neurais convolucionais .....	30
3.6.2	Overfitting e Underfitting .....	32
3.6.2.1	Dropout .....	32
3.6.2.2	Data augmentation .....	33
<b>4</b>	<b>MÉTODO PROPOSTO .....</b>	<b>35</b>
<b>4.1</b>	<b>Aquisições de imagem .....</b>	<b>35</b>
<b>4.2</b>	<b>Pré-processamento .....</b>	<b>36</b>
<b>4.3</b>	<b>Arquitetura da rede neural .....</b>	<b>37</b>
<b>4.4</b>	<b>Métricas de avaliação .....</b>	<b>38</b>
<b>5</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>40</b>
<b>5.2</b>	<b>Extração da área de interesse em novas imagens .....</b>	<b>41</b>
<b>5.3</b>	<b>Avaliação da rede neural artificial .....</b>	<b>42</b>
<b>5.4</b>	<b>Classificação em vídeo .....</b>	<b>45</b>
<b>5.5</b>	<b>Estudo de caso .....</b>	<b>45</b>
<b>6</b>	<b>CONCLUSÃO .....</b>	<b>49</b>
<b>6.1</b>	<b>Trabalhos Futuros .....</b>	<b>49</b>
	<b>REFERÊNCIAS .....</b>	<b>51</b>

## 1 INTRODUÇÃO

O desenvolvimento de carros inteligentes representa uma das mais promissoras e disruptivas tecnologias do século XXI (ACAYABA et al. 2012). Ainda de acordo com eles, a segurança é uma das principais preocupações neste campo, tendo em vista que falhas em sistemas automatizados podem ter consequências graves. Nesse sentido, a incorporação de técnicas de reconhecimento de sentimentos em carros inteligentes pode ser uma abordagem promissora para melhorar a segurança no trânsito.

Segundo Carvalho et al. (2018), o reconhecimento de sentimentos em carros inteligentes envolve a capacidade de detectar as emoções e intenções dos passageiros. Isso pode ser particularmente útil em situações de risco, quando os passageiros estão sob estresse ou agitação. Nesses casos, a detecção de sentimentos pode ajudar a entender as ações e reações dos passageiros, permitindo ao sistema tomar decisões mais precisas e seguras.

Além disso, a análise de sentimentos em vídeos pode contribuir para a prevenção de acidentes, uma vez que permite detectar comportamentos perigosos ou distrações dos passageiros. Como destacado por Liu et al. (2020), a detecção de distrações pode ajudar a reduzir o risco de acidentes causados por fatores humanos, como o uso de celulares ou outros dispositivos eletrônicos durante a condução.

Assim, as redes neurais convolucionais (CNNs) se destacam como uma técnica promissora para a classificação de sentimentos em vídeos de carros inteligentes. As CNNs são um tipo de rede neural artificial que podem ser treinadas para extrair características relevantes de imagens e vídeos, permitindo a detecção automática de objetos e padrões complexos.

De acordo com Lecun et al. (2015), as CNNs têm sido amplamente utilizadas em aplicações de visão computacional, como reconhecimento facial, classificação de imagens médicas e veiculares. Além disso, a aplicação de CNNs em vídeos permite a detecção de padrões temporais e a análise de sequências de imagens, o que é particularmente útil em aplicações de monitoramento de vídeo.

Assim, a proposta desta monografia é explorar a aplicação de redes neurais convolucionais para a classificação de sentimentos em vídeos de carros inteligentes.

Portanto, o problema da pesquisa define-se em “Como a inteligência artificial pode contribuir para a segurança no trânsito?”

Diante disso, com a detecção de emoções, o veículo poderia agir de forma autônoma, em momentos em que o condutor não teria reação para sair de situações de risco. Assim, o carro conseguiria frear, acelerar, desviar e manter a direção nos momentos em que a segurança no trânsito dependesse dessas ações. Além do que, em outros momentos, rotas com menos trânsito poderiam ser sugeridas e alertas poderiam ser emitidos em caso de emoções específicas, como na detecção de raiva.

Nesse sentido, serão utilizados conjuntos de dados para treinamento e validação do modelo, e espera-se contribuir para a segurança no trânsito por meio da utilização de técnicas de processamento de imagem para análise e classificação de sentimentos em vídeos.

## **1.1 Justificativa**

Desenvolver técnicas de visão computacional para análise em carros inteligentes. Espera-se, dessa forma, obter resultados promissores que possam ser aplicados na implementação de sistemas de veículos inteligentes mais seguros e eficientes. Além disso, a contribuição para a área de inteligência artificial e visão computacional é uma importante consequência deste trabalho, sendo assim, uma contribuição para futuras pesquisas e desenvolvimentos tecnológicos.

## **1.2 Objetivo geral**

Desenvolver uma metodologia automática capaz de classificar sentimentos em vídeos, por meio da análise facial, contribuindo para a segurança em carros inteligentes.

### **1.2.1 Objetivos Específicos**

- Realizar uma revisão bibliográfica sobre o reconhecimento facial, classificação de sentimentos em vídeos, construção e treinamento de CNNs;
- Fazer o pré-processamento dos dados coletados, como a normalização das imagens, segmentação de rostos e extração de características;

- Desenvolver uma rede neural convolucional capaz de identificar padrões específicos nas expressões faciais, com o objetivo de classificar as emoções presentes em cada vídeo;
- Avaliar a eficiência da rede convolucional em relação aos resultados obtidos, considerando-se métricas de avaliação de classificação como acurácia, precisão, sensibilidade e F1- score.

### 1.3 Organização do Trabalho

O trabalho restante consiste em cinco capítulos, os quais serão apresentados de forma resumida a seguir. No Capítulo 2 é mostrado os trabalhos relacionados com tema da presente pesquisa, expondo seus objetivos e resultados.

No Capítulo 3 é apresentado a fundamentação teórica necessária para a compreensão da metodologia proposta. Neste capítulo são descritos conceitos básicos sobre detecção de faces, reconhecimento de emoções, identificação de sentimentos por expressões faciais, reconhecimento de sentimentos em carros inteligentes, imagens digitais, Inteligência Artificial, aprendizado de máquina profundo, rede neural artificial, redes neurais convolucionais, *overfitting* e *underfitting*, *dropout* e *data augmentation*.

No Capítulo 4 é descrita a metodologia para classificar as emoções, aquisição de imagens, o pré-processamento feito, arquitetura da rede neural e métricas de avaliação.

No Capítulo 5 é exposto e discutido os resultados alcançados pela metodologia.

O Capítulo 6 mostra a conclusão deste trabalho, expondo o seu desempenho e sugestões para trabalhos futuros.

## 2 TRABALHOS RELACIONADOS

Este capítulo tem como objetivo apresentar alguns trabalhos relacionados à classificação de sentimentos em vídeos usando redes convolucionais para a segurança em carros inteligentes, como esse assunto está sendo discutido no meio tecnológico esocial e quais os impactos apresentados.

O trabalho proposto por Goodfellow et. al (2013) discute o desafio de reconhecimento de expressões faciais, que consistia em desenvolver um sistema capaz de identificar emoções em fotos de rostos humanos no conjunto de dados utilizado, chamado FER-2013. A precisão média alcançada pelos participantes ficou em torno de 68%, com equipes vencedoras utilizando redes neurais convolucionais treinadas de forma discriminativa. Além disso, algumas equipes obtiveram resultados comparáveis com métodos tradicionais, demonstrando o potencial do conjunto de dados e a importância do desafio como uma oportunidade para comparar diferentes abordagens de aprendizado de máquina no reconhecimento de expressões faciais.

Em Oliveira (2013) é apresentado um sistema computacional capaz de categorizar as emoções básicas (raiva, medo, repulsa, surpresa, alegria e tristeza) a partir das expressões faciais do usuário, as quais são captadas por uma webcam. A classificação das emoções é realizada por meio de uma rede neural artificial, que utiliza os dados extraídos de sequências de imagens em tempo real. Os resultados dos experimentos demonstram que o sistema alcançou uma taxa de reconhecimento de 63,33%, sendo que a rede neural, quando avaliada isoladamente, obteve uma taxa de reconhecimento de 89,87% nos testes.

Em Matos (2017) é estudado a identificação de expressões faciais e suas emoções correspondentes em imagens. Além de validar o método proposto, a pesquisa disponibiliza uma interface de acesso público que fornece informações sobre as emoções detectadas em uma imagem. Para embasar o trabalho, são utilizados estudos comportamentais, como a análise de Paul Ekman e as definições de emoções universais. O desenvolvimento do software envolveu técnicas de visão computacional e aprendizagem de máquina com redes neurais convolucionais. O objetivo é explorar a possibilidade de leituras em tempo real, fornecendo dados relevantes em situações em que o feedback pode ser crucial para a tomada de decisões. Mesmo com um conjunto de dados de treinamento reduzido, os resultados

mostraram uma acurácia média de teste de 65%.

Em Costa e Pedrosa (2022) o objetivo do estudo é a classificação de sentimentos em imagens digitais, uma área de pesquisa em inteligência artificial. Duas abordagens foram comparadas: a análise visual do conteúdo das imagens e a extração de legendas para descrever o conteúdo. Ambas as abordagens foram implementadas utilizando redes neurais artificiais. Os resultados mostraram que a abordagem visual teve um desempenho melhor, com 63% de acurácia na classificação de sentimentos positivos e negativos, em comparação com os 52% da abordagem baseada em transcrição de legendas.

Em vários trabalhos há uma quantidade de imagens relativamente pequena, o que influencia diretamente na taxa média de acurácia. Com isso, pode-se observar a capacidade de melhoria para esta métrica de avaliação, junto com um banco de dados maior.

### 3 REFERENCIAL TEÓRICO

Este capítulo apresenta a fundamentação teórica utilizada no desenvolvimento deste trabalho e necessária para compreensão das técnicas utilizadas no método proposto para classificação de sentimentos em vídeos.

#### 3.1 Detecção de faces

A detecção facial é uma área importante da visão computacional que envolve a identificação e localização de rostos em imagens ou vídeo, segundo Lopes (2005). Ele também afirma que é uma tarefa fundamental em muitas aplicações, como reconhecimento facial, análise de expressões faciais, detecção de idade e gênero, entre outras.

Ainda segundo o autor anteriormente citado a detecção facial é frequentemente acompanhada pela extração de características faciais. Isso envolve a identificação e localização de pontos-chave na face, como os olhos, nariz, boca e queixo. Esses pontos-chave podem ser usados para realizar análises mais detalhadas, como reconhecimento de expressões faciais e análise de emoções.

Além disso, Negri et. al (2020) afirma que a detecção facial pode enfrentar desafios, como variações na iluminação, oclusões parciais, poses diferentes e variações na aparência facial. Para lidar com esses desafios, os algoritmos de detecção facial são frequentemente aprimorados e otimizados, utilizando técnicas como pré-processamento de imagens, filtragem de ruído, normalização de iluminação e uso de modelos estatísticos.

##### 3.1.1 Reconhecimento de emoções

O reconhecimento facial e o reconhecimento de emoções são áreas de pesquisa importantes em psicologia e neurociência, e têm aplicações práticas em várias áreas, como segurança, marketing e saúde. De acordo com (Ekman e Friesen 1971), "a face humana é um canal primário para a expressão de emoções"(p. 56). O reconhecimento facial, portanto, refere-se à habilidade de reconhecer características faciais específicas, como expressões emocionais, para identificar um indivíduo.

A pesquisa sobre reconhecimento facial tem sido conduzida por muitos

anos. No entanto, a tecnologia atual permitiu o desenvolvimento de sistemas automatizados de reconhecimento facial que podem ser usados em várias aplicações práticas, como segurança, autenticação de identidade e reconhecimento de emoções. Por exemplo, em um estudo recente, Huang, Guo e Zhang (2019) desenvolveram um sistema de reconhecimento facial que usa técnicas de aprendizado de máquina para identificar indivíduos com alta precisão.

Além do reconhecimento facial, o reconhecimento de emoções também é uma área importante de pesquisa. Ekman e Friesen (1975) afirmaram que existem seis emoções básicas - felicidade, tristeza, raiva, medo, surpresa e repugnância - que são expressas de forma universal através de expressões faciais. Pesquisas recentes mostram que a capacidade de reconhecer e interpretar expressões faciais é importante para a comunicação social e pode afetar a tomada de decisões e a interação social (ADOLPHS, 2002).

O reconhecimento de emoções também tem implicações práticas em várias áreas. Por exemplo, em um estudo de 2014, Baskin-Sommers e Krusemark mostraram que a habilidade de reconhecer emoções faciais afetou a capacidade de tomada de decisões éticas em indivíduos. Além disso, o reconhecimento de emoções também pode ser usado em marketing para entender melhor as preferências do consumidor e para melhorar a experiência do usuário em produtos e serviços (TREIBLMAIER, 2018).

Embora a inteligência artificial seja frequentemente mencionada em relação ao reconhecimento facial e de emoções, é importante notar que esses processos também envolvem habilidades cognitivas e perceptivas humanas como descrito por Ekman e Friesen (1975). A habilidade de reconhecer expressões faciais e interpretar emoções é parte integrante da comunicação social humana e é essencial para a interação social bem sucedida.

### **3.2 Identificação de sentimentos por expressões faciais**

Segundo o livro "A Linguagem das Emoções" de Paul Ekman, que é um dos principais estudiosos da relação entre emoções e expressões faciais, as emoções são universais e são expressas de forma semelhante em diferentes culturas. Ekman é conhecido por ter desenvolvido a técnica do *Facial Action Coding System* (FACS), que é amplamente utilizada na psicologia e nas neurociências para a identificação e

codificação dos movimentos musculares faciais envolvidos na expressão de emoções. Ele também discute as diferentes expressões faciais associadas a diferentes emoções, bem como as diferenças entre expressões genuínas e expressões falsas.

Ekman explica que a identificação de emoções por meio de expressões faciais é uma habilidade inata em seres humanos, mas que pode ser aprimorada com treinamento. Ele discute a importância da prática e do feedback na melhoria da capacidade de reconhecimento de emoções por expressões faciais.

O autor também aborda a relação entre emoções e expressões corporais, mostrando como as emoções podem ser expressas não apenas no rosto, mas também no corpo todo. Ele explica que a leitura de emoções por meio de expressões corporais é mais difícil do que por meio de expressões faciais, mas que também é possível desenvolver essa habilidade.

O artigo "Estudo da Identificação de Emoções Através da Inteligência Artificial", de D. A. Scardua e K Marques (2018), fala sobre a inteligência artificial na identificação de emoções por expressões faciais. Segundo ele, a identificação de emoções é uma tarefa complexa e muitas vezes subjetiva. Por isso, a utilização de algoritmos de aprendizagem de máquina pode trazer mais precisão e objetividade ao processo.

Uma das técnicas mais utilizadas é a rede neural convolucional, que é treinada para detectar as expressões faciais correspondentes a cada emoção. O artigo cita vários estudos que utilizaram essa técnica e alcançaram resultados promissores na identificação de emoções como felicidade, tristeza, raiva e medo.

No entanto, o artigo também aponta algumas limitações desses estudos, como a falta de padronização nas bases de dados utilizadas e a dificuldade em identificar emoções mais sutis ou mistas. Além disso, é importante lembrar que as expressões faciais não são o único indicador de emoções e que outros fatores, como o contexto e o comportamento corporal, também devem ser considerados.

Apesar dessas limitações, a utilização de inteligência artificial na identificação de emoções por expressões faciais apresenta um grande potencial em diversas áreas, como a saúde mental, a psicologia e a segurança.

### 3.3 Tipos de emoções

Os sentimentos são uma parte fundamental da experiência humana e desempenham um papel importante em nossas interações sociais, bem como em nossa saúde emocional. A emoção é uma resposta complexa a eventos ou situações que envolvem alterações fisiológicas, cognitivas e comportamentais. Neste texto, abordou-se os seis principais sentimentos humanos: raiva, felicidade, nojo, surpresa, medo, tristeza e neutro, com base no livro "A Linguagem das Emoções", de Paul Ekman, que apresenta uma análise detalhada sobre como cada um desses sentimentos funciona.

Os sentimentos são uma parte essencial da experiência humana, e são uma das principais formas que temos de entender e reagir ao mundo ao nosso redor. Entre os sentimentos mais comuns estão a raiva, felicidade, nojo, surpresa, medo e tristeza. Cada um desses sentimentos tem uma função única e pode ser desencadeado por diferentes estímulos.

Segundo o autor, os sentimentos são desencadeados por eventos que têm uma importância para o indivíduo, seja por ser algo que ele considera uma ameaça ou uma oportunidade. Esses eventos podem ser internos ou externos e podem ser percebidos através dos sentidos ou de pensamentos. Quando um evento é percebido como importante, o cérebro é ativado para produzir uma resposta emocional.

Diante disto, as definições das emoções raiva, alegria, nojo, surpresa, medo, tristeza e neutralidade serão descritas e possuem embasamento no livro "A Linguagem das Emoções", de Paul Ekman, visto que ele oferece uma análise detalhada sobre os principais sentimentos humanos.

A raiva é um sentimento de indignação que surge quando sentimos que nossos direitos foram violados. O componente cognitivo da raiva é a percepção de injustiça, enquanto o componente fisiológico é a resposta do corpo, como aumento da frequência cardíaca e da pressão arterial. O comportamento associado à raiva inclui expressões faciais como franzir a testa, cerrar os dentes e apertar os punhos. Além disso, a raiva pode levar a comportamentos agressivos, como bater, gritar ou até mesmo atacar.

A felicidade é um sentimento positivo que é ativado quando estamos diante de uma situação agradável ou quando alcançamos um objetivo desejado.

Ekman afirma que a felicidade é caracterizada por expressões faciais que incluem o sorriso, as bochechas levantadas e os olhos brilhantes. Além disso, o autor destaca que a felicidade está associada a uma série de mudanças fisiológicas, como o aumento da frequência cardíaca e a liberação de endorfinas, que produzem uma sensação de prazer e bem-estar.

O nojo é um sentimento aversivo que é ativado quando estamos diante de uma situação desagradável, como cheiros ruins ou alimentos estragados. Ekman afirma que o nojo é caracterizado por expressões faciais que incluem o nariz enrugado e a boca apertada. Além disso, o autor destaca que o nojo está associado a uma série de reações fisiológicas, como o aumento da atividade do sistema nervoso simpático e a redução da atividade do sistema nervoso parassimpático, o que produz uma sensação de mal-estar e desconforto.

A surpresa é um sentimento que surge diante de um evento inesperado ou incomum. Segundo Ekman, a expressão facial típica da surpresa envolve uma elevação das sobrancelhas, abertura dos olhos e boca entreaberta. Essa expressão é universal e pode ser observada em pessoas de diferentes culturas e idades.

O medo é um sentimento que surge diante de uma ameaça real ou percebida. A expressão facial típica do medo envolve uma elevação das sobrancelhas, uma abertura dos olhos, um afastamento da cabeça e uma expressão de tensão muscular. Segundo Ekman, a expressão facial do medo também é universal e pode ser observada em pessoas de diferentes culturas.

A tristeza é um sentimento que ocorre quando uma pessoa experimenta uma perda significativa ou uma decepção em relação a um objetivo importante. Isso pode incluir a perda de um ente querido, um emprego, um relacionamento ou uma oportunidade desejada. A tristeza é frequentemente acompanhada por uma sensação de desesperança, solidão e desamparo. Além disso, a tristeza pode manifestar-se através de uma diminuição da energia, apetite e interesse nas atividades diárias.

A neutralidade, por outro lado, é um estado emocional em que uma pessoa não experimenta emoções intensas ou fortes. Isso pode ocorrer em situações em que não há um estímulo emocional significativo ou quando uma pessoa suprime suas emoções. Em muitos casos, a neutralidade é uma resposta emocional adaptativa e útil. Isso permite que as pessoas mantenham a calma em situações estressantes ou

tomem decisões racionais sem serem influenciadas por emoções fortes.

Em resumo, os sentimentos são uma parte fundamental da experiência humana, e cada um deles desempenha uma função única em nossa vida emocional e comportamental. O livro "A Linguagem das Emoções", de Paul Ekman oferece uma análise detalhada sobre os principais sentimentos humanos. Cada um deles é desencadeado por eventos que têm uma importância para o indivíduo, seja por ser algo que ele considera uma ameaça ou uma oportunidade. Compreender a natureza e a função de nossos sentimentos pode nos ajudar a lidar melhor com as situações desafiadoras e a promover nossa saúde emocional.

Além disso, é importante reconhecer que não existem apenas as emoções descritas, existe uma ampla gama de estados emocionais humanos, incluindo sentimentos como amor, gratidão, vergonha e muitos outros. No entanto, é válido ressaltar que apenas as emoções previamente mencionadas foram consideradas e utilizadas no escopo do projeto em análise.

### **3.4 Reconhecimento de sentimentos em carros Inteligentes**

Um estudo realizado por pesquisadores do MIT (Instituto de Tecnologia de Massachusetts) aborda justamente esse tema, investigando a aplicação de técnicas de reconhecimento de emoções em veículos autônomos. O estudo destaca que a capacidade de perceber e interpretar as emoções dos passageiros é uma habilidade importante para garantir a segurança e o conforto durante a viagem. Por exemplo, se um passageiro estiver se sentindo ansioso ou estressado, o veículo poderia ajustar a velocidade ou escolher rotas mais tranquilas para ajudar a acalmar a pessoa.

O estudo enfatiza que, à medida que os veículos inteligentes se tornam mais comuns nas estradas, é essencial que eles possam reconhecer as emoções das pessoas ao seu redor. Isso inclui a capacidade de identificar sinais de agressão, medo, ansiedade e outros estados emocionais. Ao fazer isso, os veículos podem ajustar seu comportamento de acordo com as circunstâncias, reduzindo o risco de acidentes e garantindo a segurança dos ocupantes e de outras pessoas na estrada.

No entanto, como destacado pelos autores, o reconhecimento de emoções em veículos autônomos e inteligentes é uma área de pesquisa ainda em desenvolvimento. Existem muitos desafios técnicos a serem superados para que essa

tecnologia possa ser implementada com segurança e eficácia. Um desses desafios é o desenvolvimento de sensores capazes de detectar emoções de maneira precisa e confiável. Outro desafio é a necessidade de treinar algoritmos de reconhecimento de emoções com grandes conjuntos de dados de alta qualidade. Esses dados devem ser coletados em diferentes contextos e situações, para que os algoritmos possam aprender a reconhecer as emoções de maneira confiável em uma ampla variedade de cenários.

Uma das técnicas utilizadas pelos pesquisadores é o reconhecimento de expressões faciais, que consiste em analisar as mudanças no rosto dos passageiros para identificar emoções como alegria, tristeza, raiva, surpresa e medo. Segundo o estudo, essa abordagem tem se mostrado bastante eficaz em ambientes controlados, como laboratórios, mas ainda precisa ser aprimorada para funcionar em condições mais desafiadoras, como ruídos de fundo e iluminação variável.

Diante do exposto, a capacidade de reconhecimento de emoções em veículos autônomos e inteligentes é uma área de pesquisa promissora, mas ainda em desenvolvimento. Como destacado pelos pesquisadores do MIT, essa habilidade é importante para garantir a segurança e o conforto dos passageiros, bem como dos demais usuários da estrada. No entanto, existem muitos desafios técnicos a serem superados, como a necessidade de sensores precisos e algoritmos de reconhecimento de emoções treinados em conjuntos de dados de alta qualidade. Além disso, é importante considerar a ética em torno do reconhecimento de emoções, como a privacidade dos passageiros e o potencial uso indevido dessas informações. Com mais pesquisas e avanços tecnológicos, é possível que em um futuro próximo os carros inteligentes sejam capazes de entender as emoções dos passageiros e responder adequadamente, proporcionando uma experiência mais segura e confortável na estrada.

### **3.5 Imagens digitais**

De acordo com Gonzales e Woods (2000), as imagens digitais são representações visuais de objetos, cenas ou fenômenos que são capturados, armazenados e processados em formato digital. Elas desempenham um papel fundamental em diversas áreas, como fotografia, cinema, medicina, ciência, design gráfico e muitas outras.

Além disso, os autores dizem que uma imagem digital é composta por pixels, que são os elementos básicos de uma imagem. Cada pixel contém informações sobre cor e intensidade luminosa. A resolução de uma imagem digital refere-se à quantidade de pixels que compõem a imagem e determina sua qualidade e nitidez. Quanto maior a resolução, mais detalhes podem ser capturados e exibidos na imagem.

Ainda segundo eles, o processamento de imagens digitais envolve a aplicação de técnicas e algoritmos para modificar, melhorar ou extrair informações das imagens. Essas técnicas podem incluir operações simples, como ajuste de brilho, contraste e nitidez, bem como processos mais avançados, como filtragem, segmentação, reconhecimento de padrões e reconstrução tridimensional.

Uma das áreas de aplicação mais relevantes das imagens digitais é a visão computacional, que se refere à capacidade dos computadores de analisar e compreender informações visuais. Isso inclui tarefas como detecção de objetos, reconhecimento facial, rastreamento de movimento, entre outras. A visão computacional utiliza algoritmos e técnicas de processamento de imagens para extrair informações úteis das imagens e interpretá-las.

### **3.6 Inteligência Artificial**

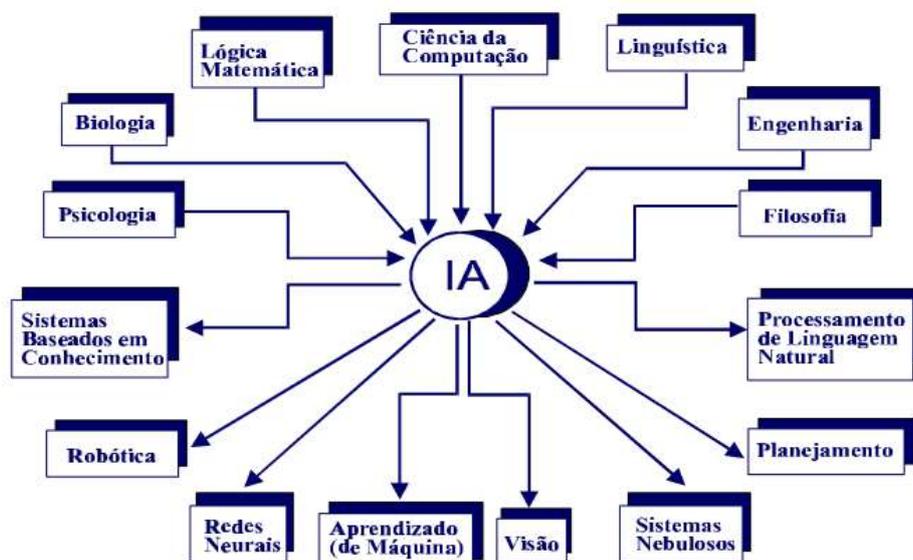
Tem-se como Inteligência Artificial (IA) a confecção de máquinas com capacidade de aprender sendo estas programadas previamente, fazendo uso de algoritmos bem elaborados e complexos que proporcionem a tomada de decisões, especulações e até interações baseadas nos dados fornecidos (DAMACENO, S. e VASCONCELOS, R., 2018). Sendo assim, pode-se entender que a inteligência artificial é uma área da ciência da computação que se dedica ao desenvolvimento de sistemas computacionais capazes de realizar tarefas que, em geral, requerem inteligência humana para serem executadas. Dentre essas tarefas estão o reconhecimento de padrões em imagens, a tomada de decisões baseada em dados, a compreensão e produção de linguagem natural, entre outras.

Ainda segundo os autores anteriormente citados, a IA não precisa necessariamente interagir com pessoas, ou simular ações humanas. O seu real conceito e objetivo é executar tarefas de forma inteligente. Desta forma, ela cria sistemas que possam aprender e melhorar suas habilidades com o tempo. Isso é feito através de algoritmos complexos que analisam grandes conjuntos de dados e

identificam padrões e correlações entre eles. Esses algoritmos são capazes de reconhecer e classificar dados com uma precisão cada vez maior à medida que recebem mais informações.

Além disso, segundo Gomes (2010), devido à amplitude de seu escopo, a inteligência artificial se relaciona com diversas outras áreas científicas, como psicologia, biologia, lógica matemática, linguística, engenharia e filosofia, conforme a Figura 1.

Figura 1 – Áreas relacionadas com a IA



Fonte: MONARD; BARANAUKAS (2000, p. 2)

Ainda de acordo com Damaceno e Vasconcelos (2000), é possível desdobrar a IA em camadas ou partes que a constituem, introduzindo assim os conceitos de Aprendizado de Máquina e Aprendizado profundo.

Dessa forma, Aprendizado de Máquina descreve a capacidade de sistemas de aprender a partir de dados específicos do problema para automatizar o processo de construção de modelos analíticos e resolver tarefas associadas.” (JANIESCH, C; ZSCHECH, P; HEINRICH, K. 2021). Em outras palavras, é uma área da inteligência artificial que se concentra em desenvolver sistemas que possam aprender e melhorar a partir de dados específicos de um determinado problema. Isso significa que, em vez de programar diretamente uma solução para um problema, o sistema é treinado com exemplos de dados e, com o tempo, ele aprende a identificar padrões e a tomar decisões com base nesses padrões.

Já o Aprendizado Profundo “é um conceito de aprendizado de máquina baseado em redes neurais artificiais. Para muitas aplicações, modelos de aprendizado profundo superam modelos rasos de aprendizado de máquina e abordagens tradicionais de análise de dados.” (JANIESCH, C; ZSCHECH, P; HEINRICH, K. 2021). Assim, é uma técnica que é capaz de aprender a partir de grandes conjuntos de dados, detectando padrões e fazendo previsões com alta precisão em várias áreas, incluindo visão computacional, processamento de linguagem natural, reconhecimento de voz, entre outras.

### 3.6.1 Aprendizado de máquina profundo

O *Deep Learning* (DL), ou Aprendizagem Profunda, “é uma subárea de Aprendizagem de Máquina que investiga técnicas para simular o comportamento do cérebro humano em tarefas como reconhecimento visual, reconhecimento de fala e processamento de linguagem natural.” (BEZERRA, E. 2016). Em suma, o autor afirma que o DL se concentra em técnicas de aprendizado de máquina que buscam simular o comportamento do cérebro humano em tarefas específicas, como reconhecimento visual, reconhecimento de fala e processamento de linguagem natural. Essa abordagem é baseada em redes neurais artificiais, que são projetadas para se assemelhar às redes neurais do cérebro humano. A ideia é que, ao imitar o modo como o cérebro humano processa e aprende informações, os sistemas de IA possam ser mais eficazes em tarefas que exigem alto nível de complexidade e precisão, como reconhecimento de imagens e tradução de idiomas.

#### 3.6.1.1 Rede neural artificial

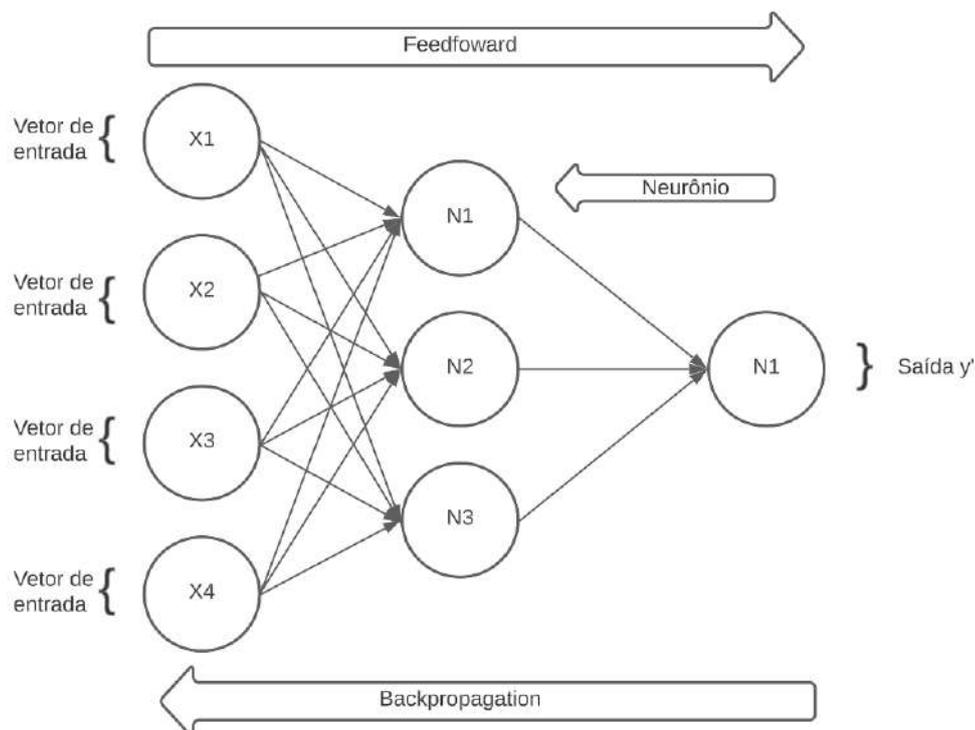
Ainda segundo o autor citado anteriormente, “A arquitetura de uma rede neural diz respeito ao modo pelo qual suas unidades de processamento estão conectadas. Essa arquitetura influencia diretamente nos problemas que uma Rede Neural Artificial (RNA) pode resolver e na forma de realizar seu treinamento.” Dessa forma, arquitetura de uma rede neural se refere à maneira como as suas unidades de processamento estão conectadas.

Isso significa que a forma como as diferentes unidades (também conhecidas como neurônios) estão interconectadas determina como a rede neural irá funcionar e quais tipos de problemas ela poderá resolver. Além disso, a arquitetura também influencia diretamente a forma como a rede neural é treinada, ou seja, como

ela aprende a realizar uma determinada tarefa. Portanto, é importante escolher a arquitetura adequada para cada tipo de problema a fim de obter melhores resultados.

As redes neurais artificiais são algoritmos de Aprendizado de Máquina que imitam o comportamento dos neurônios biológicos do cérebro humano. Esses neurônios são ativados em determinadas situações, o que leva o corpo a realizar uma ação relacionada em resposta a um estímulo. Elas podem ser compostas por uma ou múltiplas camadas de neurônios artificiais, que são unidades de processamento que simulam a atividade neural no cérebro humano, como mostra na Figura 2, o exemplo da arquitetura com uma camada de neurônios escondida.

Figura 2 – Exemplo de rede neural de múltiplas camadas



Fonte: Adaptador de Lauande 2022

As redes neurais artificiais são compostas por uma ou múltiplas camadas de neurônios artificiais que estão interligados e alimentados por funções de ativação que auxiliam a ativá-los ou desativá-los. Assim como os algoritmos de máquinas convencionais, essas redes aprendem valores específicos durante a fase de treinamento.

Existem diversas funções de ativação disponíveis, cada uma adequada para diferentes tipos de problemas. No método proposto neste trabalho, foram

utilizadas as funções *Softmax* e *ReLU*.

A função de ativação *Softmax* é amplamente utilizada em problemas de classificação multiclasse. Ela converte a saída de um modelo de aprendizado de máquina em uma distribuição de probabilidades, atribuindo-as a cada classe possível. Esta função garante que as probabilidades atribuídas a cada classe estejam sempre no intervalo  $[0, 1]$  e que a soma de todas as probabilidades seja igual a 1. Isso facilita a interpretação dos resultados e permite a tomada de decisões. Sua fórmula está presente na Equação 1.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \text{ for } j = 1, \dots, K \quad (01)$$

A função de ativação *ReLU* (*Rectified Linear Unit*) é uma função amplamente utilizada em redes neurais e outros modelos de aprendizado de máquina. Sua principal característica é a não-linearidade, permitindo a modelagem de relações complexas entre os dados de entrada e a saída do modelo.

A função *ReLU* define que se o valor de entrada for positivo, a saída será igual ao próprio valor de entrada. Caso contrário, se o valor de entrada for negativo, a saída será zero. Em termos matemáticos, a função *ReLU* pode ser representada como  $f(x) = \max(0, x)$ , onde  $x$  é o valor de entrada e  $f(x)$  é o valor de saída, conforme a Equação 2.

$$f(x) = \max(0, x) \quad (02)$$

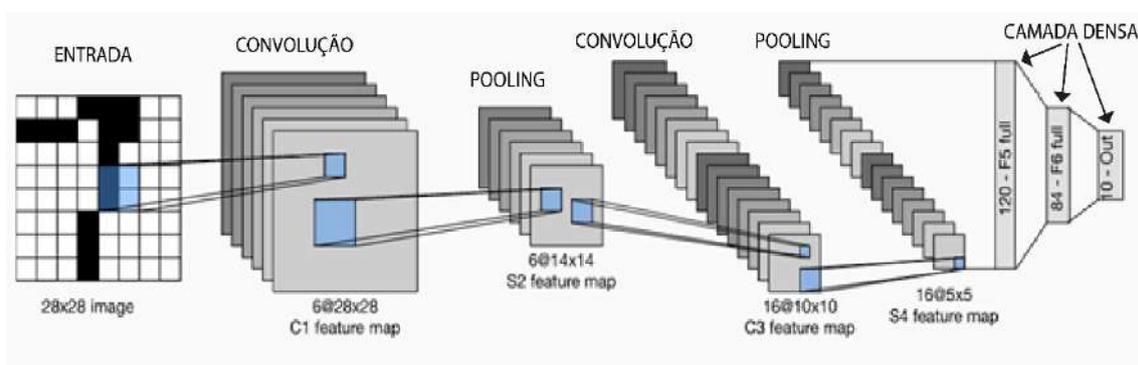
Após a predição do valor na última camada do *feedforward*, começa o cálculo da função de custo (função *loss*). Durante o *backpropagation*, o erro entre a saída esperada e a saída gerada pela rede é calculado e propagado de volta para a rede neural. Esse erro é usado para ajustar os pesos das conexões para que a próxima saída gerada pela rede esteja mais próxima da saída esperada. O processo de ajuste dos pesos é realizado utilizando um algoritmo de otimização como o gradiente descendente. O gradiente descendente é um método iterativo que busca encontrar o mínimo global de uma função de perda. Nesse caso, a função de perda é a diferença entre a saída gerada pela rede e a saída esperada.

O algoritmo de *backpropagation* é importante porque permite que as redes neurais artificiais aprendam com exemplos de treinamento e possam generalizar para novos exemplos. Isso é possível porque a rede é capaz de ajustar seus pesos para reconhecer padrões nas entradas e gerar a saída esperada. É importante ressaltar que a eficiência do treinamento depende da escolha adequada dos parâmetros da rede neural, como o número de neurônios, camadas, função de ativação e taxa de aprendizagem. Além disso, é importante que os exemplos de treinamento sejam representativos e variados o suficiente para que a rede neural seja capaz de generalizar para novos exemplos.

### 3.6.1.2 Redes neurais convolucionais

Para Zhang et al. (2021) as Redes Neurais Convolucionais (ou *Convolutional Neural Networks* – CNNs são um tipo de arquitetura de rede neural artificial comumente usada em tarefas de reconhecimento de imagem. A principal ideia por trás das CNNs é utilizar operações de convolução para extrair características relevantes das imagens de entrada, o que permite a identificação de objetos e padrões em imagens complexas. A arquitetura básica de uma CNN é composta por camadas de convolução, camadas de pooling e camadas totalmente conectadas, conforme a Figura 3.

Figura 3 – Rede convolucional – LeNet5.



Fonte: Adaptado de (ZHANG et al. 2021)

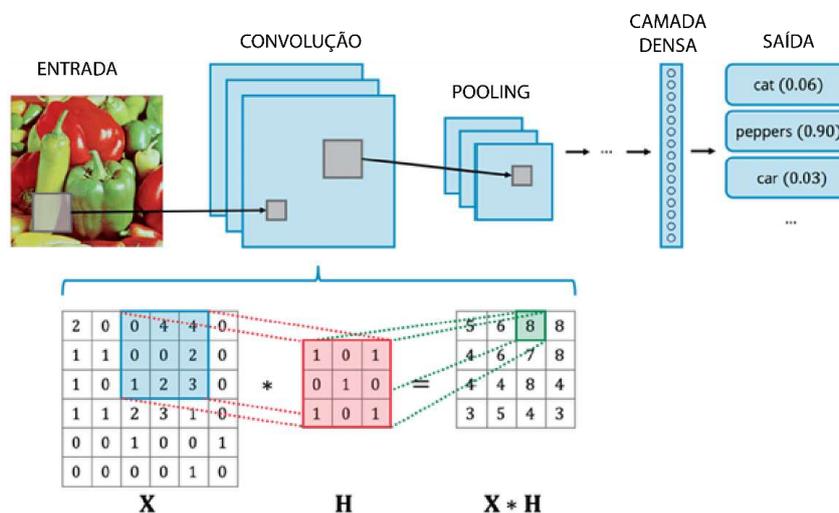
Ainda de acordo com os autores supracitados, as camadas de uma CNN são a essência do processo de convolução, que é a chave para o sucesso das redes neurais convolucionais. Existem basicamente três tipos de camadas em uma CNN: camadas de convolução, camadas de pooling e camadas completamente conectadas.

As camadas convolucionais em uma CNN são compostas por um conjunto de filtros (ou *kernels*) que deslizam sobre a imagem de entrada (Figura 4). Cada filtro é responsável por extrair uma característica específica da imagem, como bordas, linhas e texturas. Durante o processo de convolução, o filtro é multiplicado pelos pixels correspondentes da imagem e, em seguida, somado. Isso resulta em um mapa de características que destaca a presença da característica específica representada pelo filtro.

A primeira camada é a camada de convolução, que é responsável por aplicar a operação de convolução à entrada da rede. A convolução é uma operação matemática que combina a entrada com um conjunto de filtros para gerar um conjunto de mapas de característica, como citado por Passos (2021). A autora também afirma que esses mapas de características capturam a presença de certas características nas imagens, como bordas, formas, texturas, entre outros. Cada mapa de características é gerado pelo uso de um filtro específico.

Ainda por Passos (2021), a camada de *pooling* é geralmente usada após a camada de convolução para reduzir o tamanho dos mapas de características. A redução de tamanho é importante para reduzir o número de parâmetros e diminuir o tempo de processamento. A operação de *pooling* é feita aplicando uma função em cada sub-região do mapa de características, selecionando o valor máximo (*max-pooling*) ou a média (*average-pooling*), por exemplo, para representar a sub-região inteira.

Figura 4 – Exemplo da aplicação de *Kernel 3x3*



Fonte: Adaptado de (JIANG et al. 2020)

As camadas completamente conectadas são usadas para realizar a classificação final dos dados de entrada. Essas camadas são semelhantes às camadas de uma rede neural convencional, em que todas as saídas de uma camada são conectadas às entradas da próxima camada. Essas camadas têm uma grande quantidade de parâmetros, que precisam ser ajustados durante o treinamento da rede.

### 3.6.2 Overfitting e Underfitting

O *overfitting* e o *underfitting* são dois problemas que podem ocorrer quando se treina um modelo de aprendizado de máquina, incluindo as CNNs. Segundo Zhang et al. (2021), o *overfitting* ocorre quando o modelo aprende a representar muito bem os dados de treinamento, mas falha ao generalizar para novos dados que nunca foram vistos antes. Isso pode ocorrer quando o modelo se ajusta demais aos dados de treinamento, capturando até mesmo o ruído presente nesses dados, em vez de aprender os padrões subjacentes que são importantes para fazer previsões precisas.

Por outro lado, o *underfitting* ocorre quando o modelo não consegue aprender adequadamente a partir dos dados de treinamento, falhando em capturar os padrões importantes nos dados. Isso resulta em um modelo que não se ajusta bem nem mesmo aos dados de treinamento e, portanto, não pode fazer previsões precisas em novos dados (ZHANG ET AL. 2021).

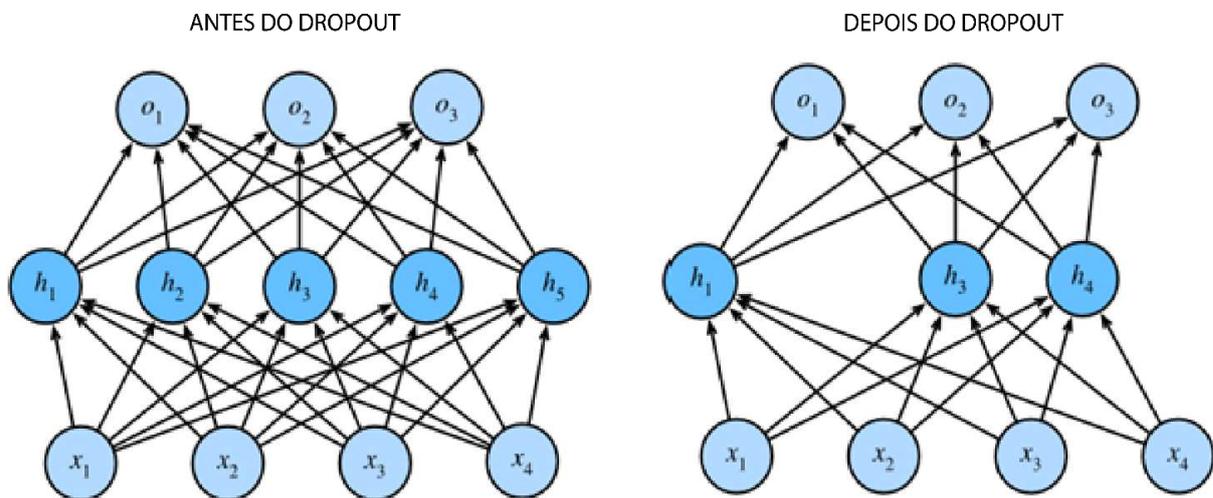
Ainda de acordo com o autor anteriormente citado, para evitar o *overfit*, técnicas como Transferência de Aprendizado, *data augmentation* e *Dropout*, podem ser utilizadas durante o treinamento para impedir que o modelo se ajuste demais aos dados de treinamento. Já para resolver o *underfit*, podem ser utilizadas técnicas como a adição de camadas ou aumentar o número de neurônios para que o modelo possa aprender melhor a partir dos dados de treinamento. É importante encontrar o equilíbrio certo entre o ajuste aos dados de treinamento e a capacidade de generalizar para novos dados, a fim de obter um modelo com bom desempenho. Dentre essas técnicas citadas, o *dropout* e *data augmentation* foram utilizadas neste trabalho.

#### 3.6.2.1 Dropout

O *Dropout* é uma técnica de regularização muito utilizada RNAs profundas,

que tem como objetivo reduzir o *overfitting* durante o treinamento da rede. A técnica consiste em remover aleatoriamente algumas unidades (neurônios) da rede neural durante o treinamento, o que força a rede a aprender características relevantes de forma mais robusta e independente de neurônios específicos. Esse processo faz com que a rede se torne menos dependente de neurônios individuais, o que pode melhorar a capacidade de generalização da rede e, assim, reduzir o *overfitting*, conforme Shorten e Khoshgoftaar (2019). Na Figura 5 contém a representação de uma rede antes e depois do *dropout*.

Figura 5 – Antes e depois da aplicação do *dropout* em uma rede neural



Fonte: Adaptado de (SHORTEN et al. 2019)

### 3.6.2.2 Data augmentation

*Data augmentation* é uma técnica amplamente utilizada em aprendizado de máquina e em redes neurais para aumentar a quantidade e variedade de dados de treinamento disponíveis, melhorando o desempenho do modelo, como dito por Shorten e Khoshgoftaar (2019).

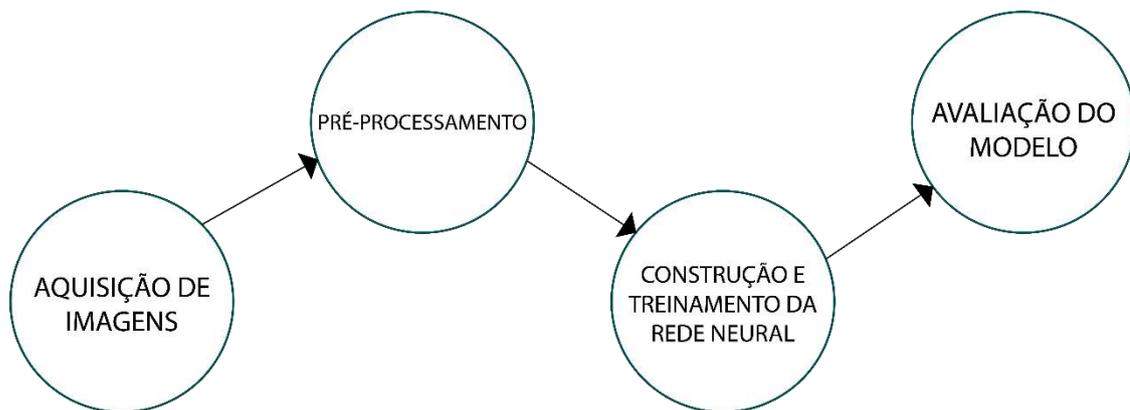
A técnica de *data augmentation* consiste em gerar novos dados de treinamento a partir dos dados existentes, aplicando transformações diversas, como rotações, deslocamentos, zoom, ajustes de brilho e contraste, entre outros. Essas transformações permitem que o modelo treine com exemplos mais variados, melhorando a capacidade de generalização e reduzindo o risco de *overfitting*.

A escolha de técnicas e algoritmos de transformação de dados depende principalmente do tipo de problema a ser resolvido, pois é importante que o cientista de dados tenha conhecimento sobre o domínio do problema para evitar que as transformações prejudiquem a rotulação das imagens. Por exemplo, ao trabalhar com imagens que representam números ou letras, é necessário ter cuidado ao utilizar transformações como rotação ou espelhamento, pois essas operações podem acabar alterando o significado da imagem, fazendo com que um número 6 se torne um número 9, por exemplo. É fundamental avaliar as técnicas de *data augmentation* de acordo com o contexto do problema e realizar experimentos para entender como as transformações afetam a qualidade dos dados e o desempenho do modelo.

## 4 MÉTODO PROPOSTO

Neste capítulo são descritas as etapas utilizadas na metodologia proposta para a classificação de sentimentos em imagens digitais. A metodologia está dividida em seis etapas. Em síntese, a primeira etapa é a aquisição das imagens que foram obtidas no Kaggle, utilizadas no desafio “*Facial Expression Recognition 2013*” (FER-2013). Na segunda etapa é realizado o pré-processamento das imagens. Na terceira, é feita a construção da rede neural. Na quarta etapa, é realizada a avaliação dela. A Figura 6 mostra o fluxo das etapas propostas de maneira resumida.

Figura 6 – Diagrama do método proposto



Fonte: Elaborado pelo autor.

### 4.1 Aquisições de imagem

Foi escolhido o conjunto de dados disponibilizado no Kaggle, especificamente o conjunto de dados FER2013. Este conjunto de dados contém imagens em escala de cinza de rostos de pessoas, cada uma classificada de acordo com uma das sete emoções básicas: raiva, nojo, medo, felicidade, tristeza, surpresa e neutralidade. Foram utilizadas cerca de 35 mil imagens para treinamento e validação do modelo.

Uma das vantagens de utilizar esse conjunto de dados foi a disponibilidade e o fácil acesso a um grande número de imagens rotuladas. Além disso, as imagens foram coletadas de diferentes origens, incluindo filmes, programas de televisão e vídeos online, o que garante uma variedade de expressões faciais e situações. No

entanto, uma desvantagem do conjunto de dados é a qualidade das imagens, que em alguns casos são baixas, com baixa resolução ou com problemas de iluminação, o que pode afetar o desempenho do modelo de reconhecimento de expressões faciais.

Em geral, a utilização do conjunto de dados FER2013 foi fundamental para a realização deste projeto de pesquisa, fornecendo uma grande quantidade de imagens rotuladas para treinamento e validação do modelo. A partir desses dados, foi possível desenvolver um modelo preciso de reconhecimento de expressões faciais, contribuindo para a área de visão computacional e inteligência artificial.

## 4.2 Pré-processamento

Durante a etapa de pré-processamento neste trabalho, é feito o *data augmentation*, utilizado para preparar conjuntos de dados de imagens para treinar e avaliar uma rede neural para classificação de emoções em imagens. O seu objetivo é realizar a normalização, rotação, *flip* horizontal e mudança de zoom em imagens para que possam ser utilizadas no treinamento da rede.

A normalização das imagens é feita por meio de um parâmetro, que divide todos os valores dos pixels das imagens por 255, que é o valor máximo que um pixel pode ter. Em seguida, a rotação é aplicada, para definir a amplitude máxima do ângulo de rotação em 7 graus. Além disso, é aplicado o *flip* horizontal nas imagens para que a rede possa aprender a reconhecer as emoções independentemente da orientação da imagem. Por fim, é aplicado um zoom aleatório nas imagens para que a rede possa aprender a reconhecer as emoções independentemente do tamanho das imagens, com a amplitude máxima do zoom em 0.2.

Após a preparação dos dados, é criado o gerador de treinamento. Nesta etapa, são definidos diversos parâmetros, como o tamanho da imagem de entrada definida pela proporção 48x48 pixels. Um parâmetro que define o número de imagens que serão carregadas em cada lote. Neste caso, o valor é definido como 16, o que significa que 16 imagens serão carregadas em cada lote. Isso é necessário porque, em muitos casos, as imagens a serem utilizadas no treinamento da rede neural são grandes demais para serem carregadas todas de uma vez na memória do computador.

Ao utilizar o gerador de treinamento, é possível carregar apenas um lote de imagens por vez, pré-processá-las e, em seguida, alimentar a rede neural com esses lotes. O próximo parâmetro a ser utilizado é o "*class mode*", definido como

“*categorical*”, pois é necessário classificar 7 tipos de sentimentos. Por fim, o último parâmetro “*shuffle*” utilizado é para determinar que os dados serão embaralhados antes de serem carregados em cada lote. Isso é importante para evitar que a rede neural decore a ordem dos dados durante o treinamento e, assim, seja capaz de generalizar melhor para dados não vistos anteriormente.

Logo após, é criado um gerador de teste e o conjunto de dados de teste utilizando os mesmos parâmetros que foram utilizados para o conjunto de treinamento, exceto pelo tamanho do lote, que agora é definido como 1 para permitir a avaliação de uma imagem por vez. O parâmetro “*shuffle*” é definido como “*False*”, para que as imagens sejam lidas na mesma ordem em que estão na pasta e assim possam ser validadas.

### 4.3 Arquitetura da rede neural

Para a construção da rede, foi escolhida uma rede neural convolucional (CNN) utilizando a biblioteca Keras. Com objetivo é realizar a classificação de imagens de expressões faciais em 7 emoções diferentes.

O primeiro passo foi definir algumas variáveis importantes: o número de detectores/filtros utilizados na rede (32), a quantidade de classes (7), as dimensões das imagens de entrada (48 x 48) e a quantidade de épocas que a rede será treinada (70).

A seguir, é criada a rede neural utilizando a classe “*Sequential*”, que permite criar camadas sequencialmente. A primeira camada é a de convolução, que utiliza 32 filtros de tamanho 3x3 e função de ativação ReLU. A entrada é definida como uma imagem de largura, altura e profundidade (RGB) de 48x48x3 pixels.

Após a camada de convolução, é adicionada uma camada de normalização (*Batch Normalization*), que tem como objetivo normalizar as ativações da camada anterior. Em seguida, há outra camada de convolução com 32 filtros de tamanho 3x3 e função de ativação ReLU, seguida de outra camada de normalização e uma camada de *max pooling* com tamanho de pool 2x2, que reduz a dimensão da imagem pela metade.

Em seguida, há uma camada de *dropout*, que tem como objetivo reduzir o *overfitting* (sobreajuste), que ocorre quando a rede se ajusta muito bem aos dados de treinamento e não generaliza bem para dados novos. Neste caso, 20% dos neurônios da camada anterior são gerados aleatoriamente durante o treinamento.

O processo se repete por mais três blocos todos com as mesmas características, com as camadas de convolução e normalização sendo duplicadas em cada bloco e o número de filtros sendo dobrado em cada bloco (64, 128 e 256).

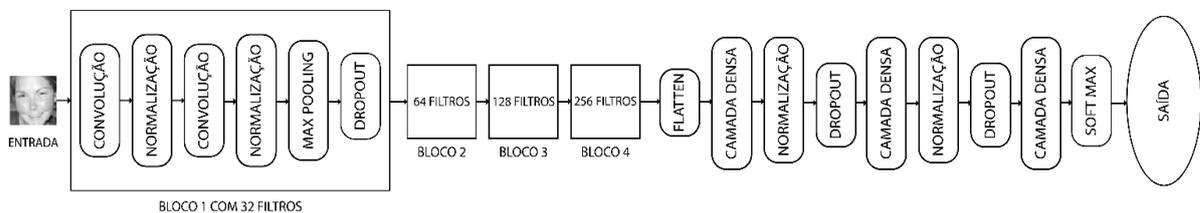
Após os quatro blocos de convolução, há uma camada de “*flatten*”, que transforma a saída das camadas convolucionais em um vetor unidimensional. Em seguida, há duas camadas densas (*fully connected*), com cada uma tendo 64 neurônios e função de ativação ReLU, seguida de uma camada de *dropout*.

A camada final é a camada de saída, com 7 neurônios e função de ativação *softmax*, que transforma as saídas em uma distribuição de probabilidade.

Por fim, é definida a compilação da rede, utilizando um otimizador, a função de perda e a acurácia. A função de perda é utilizada para medir o quão bem a rede está realizando a classificação, enquanto o otimizador é responsável por ajustar os pesos da rede para minimizar a função de perda durante o treinamento. A acurácia é utilizada para avaliar o desempenho da rede durante o treinamento e teste.

A Figura 7 – mostra o fluxo da arquitetura proposta anteriormente.

Figura 7– Fluxo da arquitetura da rede neural



Fonte: Elaborado pelo autor.

#### 4.4 Métricas de avaliação

Para avaliar o sistema proposto, os resultados dos

- Verdadeiro positivo (VP): quando o sistema identifica corretamente que uma imagem contém determinado sentimento;
- Verdadeiro negativo (VN): quando o sistema identifica corretamente que uma imagem não contém determinado sentimento;
- Falso negativo (FN): quando o sistema falha em identificar que uma imagem contém determinado sentimento;
- Falso positivo (FP): quando o sistema identifica que uma imagem contém determinado sentimento, quando na verdade não contém.

Para avaliar o desempenho do classificador, é comum utilizar o cálculo de algumas estatísticas como Precisão, Sensibilidade, F1 – Score e Acurácia.

A precisão é a métrica que indica a proporção de resultados corretos positivos em relação a todos os resultados positivos previstos pelo modelo. É definida como a razão entre o número de verdadeiros positivos e a soma dos verdadeiros positivos e falsos positivos.

$$\text{Precisão} = \frac{VP}{VP+FP} \quad (03)$$

A sensibilidade é uma métrica que indica a proporção de resultados corretos positivos em relação a todos os exemplos positivos que existem no conjunto de dados. É definida como a razão entre o número de verdadeiros positivos e a soma dos verdadeiros positivos e falsos negativos.

$$\text{Sensibilidade} = \frac{VP}{VP+FN} \quad (04)$$

O *F1-Score* é uma medida de avaliação de desempenho de um modelo que combina a precisão e a sensibilidade em uma única pontuação. Ele é calculado como a média harmônica da precisão e da sensibilidade.

$$F1\text{-Score} = 2 * \frac{\text{precisão} * \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}} \quad (05)$$

A acurácia é a métrica que indica a proporção de resultados corretos em relação a todos os exemplos no conjunto de dados. É definida como a razão entre o número de verdadeiros positivos e verdadeiros negativos e a soma de todos os resultados.

$$\text{Acurácia} = \frac{VP+VN}{VP+FN+VN+FP} \quad (06)$$

As métricas de avaliação apresentadas foram usadas para avaliar o desempenho da metodologia desenvolvida neste trabalho, considerando os sentimentos classificados corretamente como verdadeiros positivos.

## **5 RESULTADOS E DISCUSSÃO**

Neste capítulo, serão expostos e discutidos os resultados do treinamento da RNA e a sua aplicação, na qual foram feitos em fotos e vídeos de simulação com uma pessoa dirigindo um veículo em dois horários do dia: manhã e noite, para que fosse observado a influência da iluminação na leitura do vídeo, e também com o uso de óculos de grau, para saber se o seu uso interfere no resultado de classificação e extração da área de interesse.

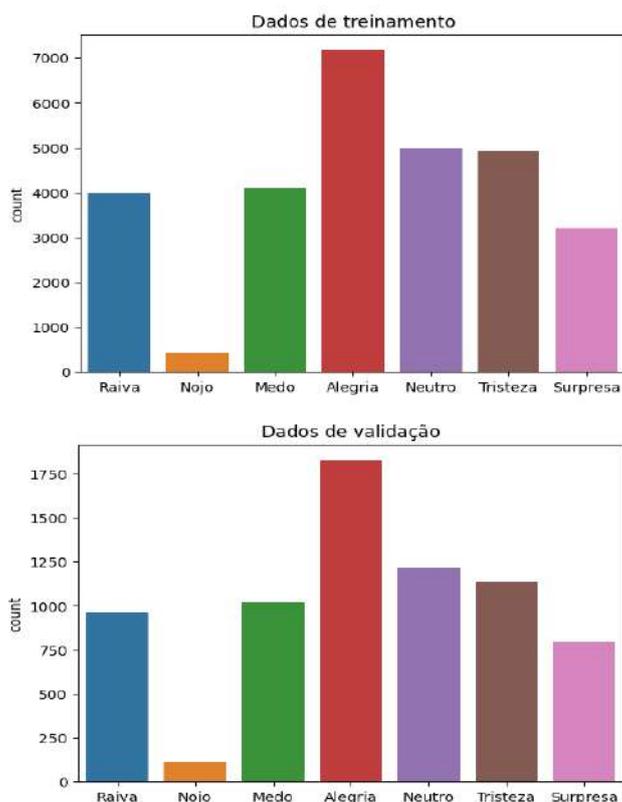
### **5.1 Imagens de treino e validação**

Após todos os pré-processamentos vistos anteriormente, foi possível obter uma quantidade significativa de imagens, sendo assim, a base de treino ficou com 3.993 imagens de raiva, 436 de nojo, 4.103 de medo, 7.164 de felicidade, 4.982 de neutro, 4.938 de tristeza e 3.205 imagens da emoção surpresa.

Para as imagens de validação a quantidade de imagens é menor, logo, a emoção de raiva ficou com 960 imagens, nojo ficou com 111, medo com 1.018, felicidade com 1.825, neutro com 1.216, tristeza com 1.139 e surpresa com 797.

Por fim, é criado dois gráficos que mostram a quantidade de classes e imagens dentro de cada uma delas, conforme a Figura 8.

Figura 8 – Quantidade de imagens nos conjuntos de treino e validação.



Fonte: Elaborado pelo autor.

## 5.2 Extração da área de interesse em novas imagens

Para a identificação de imagens com a finalidade de analisar emoções, um detector de faces desenvolvido por Pisarevsky foi empregado para que as imagens se assemelham às imagens presentes no conjunto de treinamento. Para tanto, foi implementado um sistema de reconhecimento facial em conjunto com uma rede neural convolucional (CNN) já treinada em um conjunto de dados de imagens de rostos.

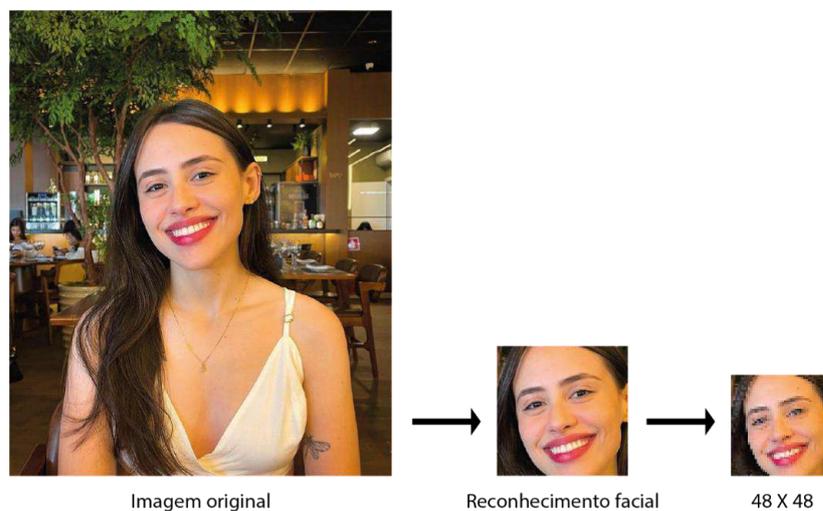
Inicialmente, a imagem é lida e sua dimensão é verificada. Em seguida, um objeto que contém as informações necessárias para a detecção de faces é criado. A partir dessas informações, uma região de interesse é estabelecida na imagem, que contém a face detectada. Essa área de interesse é redimensionada para 48x48 pixels e normalizada para valores entre 0 e 1, a fim de ser utilizada como entrada para a CNN.

Por fim, a CNN é carregada a partir de um modelo previamente treinado, e a previsão é realizada utilizando a imagem normalizada como entrada. A classe com

maior probabilidade é selecionada, e a previsão final é apresentada ao usuário. Esse processo permite a identificação de sentimentos nas imagens analisadas e é de grande importância para aplicações em áreas como reconhecimento facial, análise de emoções e detecção de expressões faciais.

Na Figura 9 é possível observar a imagem de entrada original, após a detecção da face ser feita, e antes de entrar na rede neural.

Figura 9 – Exemplo do processo de identificação de faces para a classificação em novas imagens.

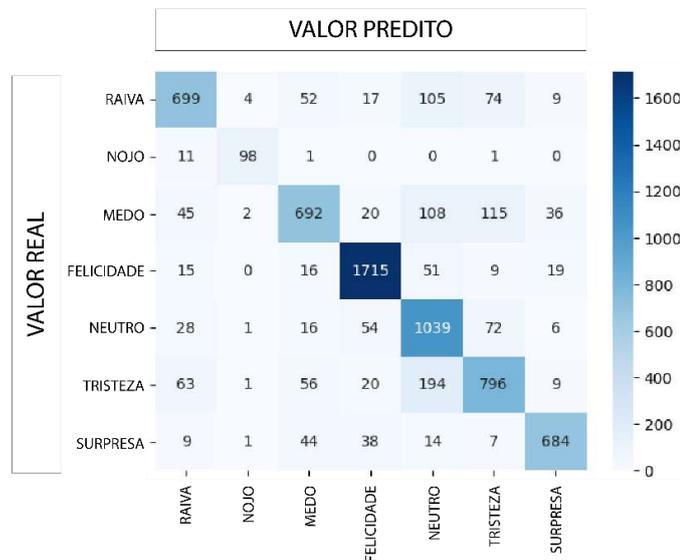


Fonte: Elaborado pelo autor.

### 5.3 Avaliação da rede neural artificial

A etapa de avaliação da RNA foi realizada para obter a matriz de confusão, que permitiu a extração dos valores de verdadeiro positivo, verdadeiro negativo, falso positivo e falso negativo. A partir disso, tornou-se viável identificar as métricas de avaliação propostas anteriormente: precisão, sensibilidade, *F1-Score* e acurácia, conforme será observado nas figuras abaixo.

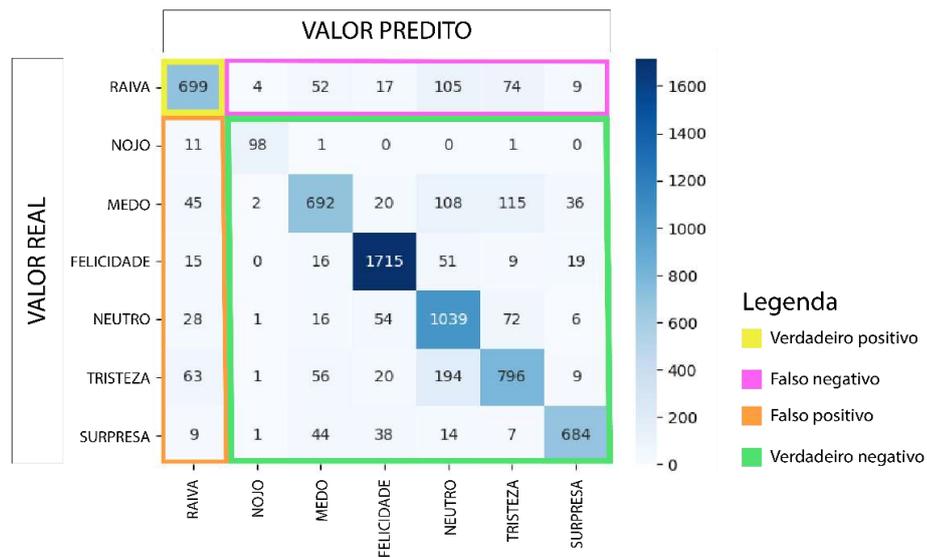
Figura 10 – Matriz de confusão



Fonte: Elaborado pelo autor.

Na figura 10, observa-se a matriz de confusão, que foi utilizada como métrica de avaliação, por permitir a visualização de valores para análise. Conforme o IBM, a Matriz de Confusão exibe a distribuição dos registros em termos de suas classes atuais e de suas classes previstas, o que indica a qualidade do modelo atual.

Figura 11 – Matriz de confusão: observação da emoção raiva



Fonte: Elaborado pelo autor.

Na figura acima, foi analisado separadamente a emoção “raiva”, identificando os valores de verdadeiro positivo (699), verdadeiro negativo, falso positivo e falso negativo, sendo eles a soma de todos os valores dentro da caixa de sua respectiva cor, conforme a legenda.

Tabela 1 – Métricas de avaliação

<b>EMOÇÕES</b>	<b>PRECISÃO</b>	<b>SENSIBILIDADE</b>	<b>F1-SCORE</b>
<b>RAIVA</b>	0.80	0.73	0.76
<b>NOJO</b>	0.92	0.88	0.90
<b>MEDO</b>	0.79	0.68	0.73
<b>FELICIDADE</b>	0.92	0.94	0.93
<b>NEUTRO</b>	0.69	0.85	0.76
<b>TRISTEZA</b>	0.74	0.70	0.72
<b>SURPRESA</b>	0.90	0.86	0.88
<b>ACURÁCIA</b>			0.81

Fonte: Elaborado pelo autor.

Na Tabela acima, observa-se os resultados das métricas propostas. A partir destes resultados, analisa-se a acurácia de 81%, o que implica dizer que o modelo foi capaz de classificar corretamente 81% do total de imagens utilizadas para a validação. Além disso, a métrica de sensibilidade apresenta a capacidade do algoritmo de identificar corretamente as imagens de determinados sentimentos, ou seja, para raiva, ele foi capaz de acertar 73%. O valor de precisão é relacionado com a capacidade de identificar assertivamente as emoções, o que diz respeito a sensibilidade, logo, na emoção “raiva”, foi possível obter 80% de precisão dentro dos 73% de imagens classificadas corretamente.

O *F1-score* une a precisão e sensibilidade em uma única métrica. Ele foi levado em consideração pela quantidade de classes propostas e a disparidade dos resultados entre elas. A partir dele, é possível medir o desempenho do modelo em uma única medida, considerando a capacidade de evitar falsos positivos e falsos negativos. Seu valor varia de 0 a 1, sendo 1 o melhor valor possível. Sendo assim, ele avalia o desempenho global do modelo proposto, avaliando tanto os seus resultados positivos, quanto negativos.

## 5.4 Classificação em vídeo

Para a classificação dos sentimentos em vídeo, primeiramente foi feita a leitura do vídeo. Em seguida é criada uma lista com as emoções que serão detectadas no vídeo. Desta forma, é iniciado um *loop* para percorrer todos os frames do vídeo. Dentro desse *loop*, é feita a leitura de cada *frame*, e é realizada a detecção de faces em cada um deles.

Caso sejam encontradas faces, são desenhados retângulos ao redor de cada uma delas. Em seguida, é extraída uma região de interesse (ROI) para cada uma das faces encontradas, utilizando as coordenadas dos retângulos desenhados. A área de interesse é redimensionada para um tamanho de 48x48 pixels e normalizada para valores entre 0 e 1.

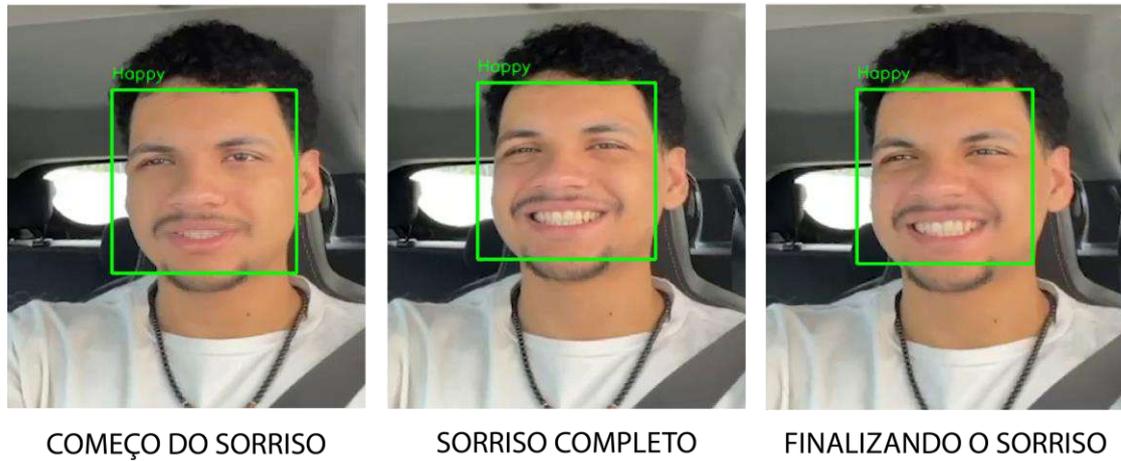
Após essa etapa, é realizada a classificação da emoção presente na ROI utilizando o modelo treinado. A previsão é então utilizada para determinar a emoção presente na área de interesse e é desenhado um texto na imagem com a emoção detectada. Por fim, é exibido o *frame* processado e escrito no vídeo de saída. Esse processo é repetido até que todos os *frames* do vídeo sejam processados.

## 5.5 Estudo de caso

Foi analisado o desempenho da RNA em vídeos dentro de um carro, com a intenção de simular uma pessoa dirigindo. A rede foi testada em 4 tipos diferentes de ambientes, sendo eles: motorista sem óculos e com óculos dirigindo de dia e de noite. Além disso, para o vídeo, foi utilizado uma câmera de 12 megapixel, com abertura de lente 1.6, gravando em qualidade HD a 30 *frames* por segundo.

Ao analisar o vídeo durante o dia, foi perceptível a facilidade de identificação da face e das emoções, dando destaque para a felicidade, pois a sua classificação foi possível mesmo antes de um sorriso estar formado por completo, e até o momento em que ele finaliza, conforme a figura 12.

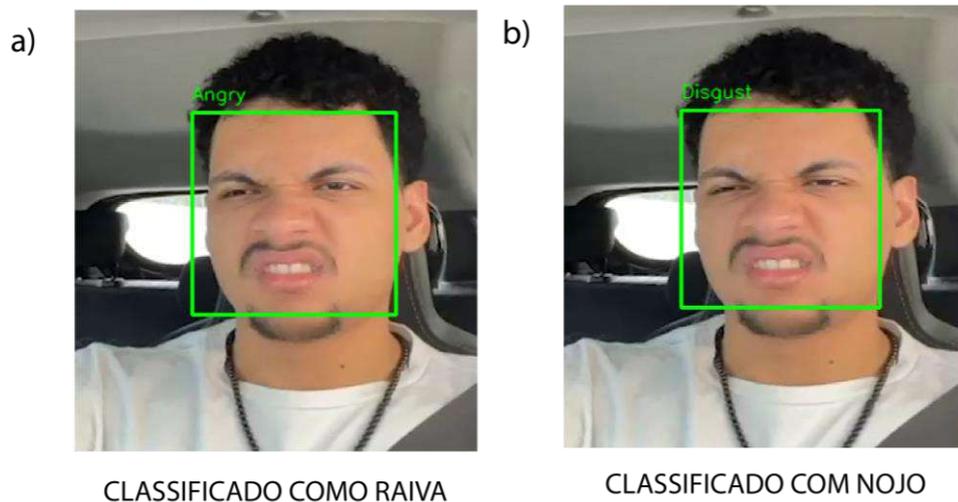
Figura 12 – Classificação em diferentes etapas do sorriso



Fonte: Elaborado pelo autor

Além disso, em emoções nas quais o algoritmo não teve um desempenho tão bom, notou-se uma certa confusão para classificação de raiva na imagem “a” e nojo na imagem “b”, fazendo com que ocorressem classificações distintas em *frames* seguidos do vídeo, conforme a Figura 13.

Figura 13 – Classificação diferente em imagens semelhantes



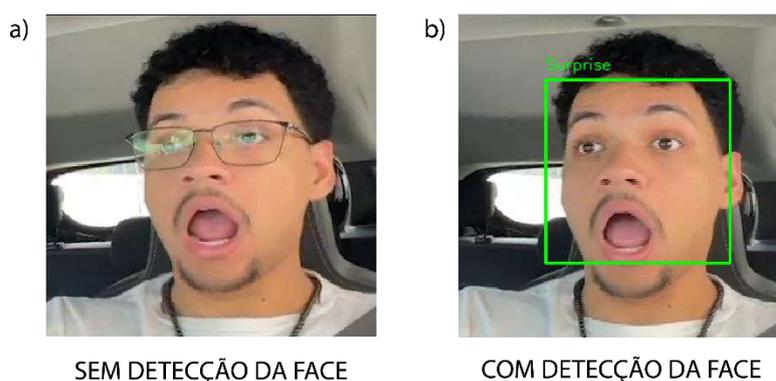
Fonte: Elaborado pelo autor.

O algoritmo teve dificuldade para esta classificação pois ambas as emoções anteriormente citadas possuem características semelhantes na imagem analisada, sendo elas, sobrancelhas franzidas, fazendo com que apareçam rugas na

testa. Outra semelhança é a tensão muscular na face, contraindo muito os mesmos músculos entres as duas emoções. Além disso, o olhar pode se assemelhar muito de pessoa para pessoa, visto que há um aumento na intensidade do olhar.

No vídeo onde foi testado com óculos, observou-se uma capacidade inferior do algoritmo de identificar a face para que ocorresse a classificação, mostrado na imagem “a” da Figura 14. Tal fato ocorre pelo motivo dos óculos poderem alterar a forma e proporção do rosto para a máquina, assim dificultando a identificação da face e extração da área de interesse, além de refletir a luz, fazendo com que a região dos olhos e suas características não sejam identificadas, como é mostrado na Figura 14 durante a detecção da emoção “surpresa”.

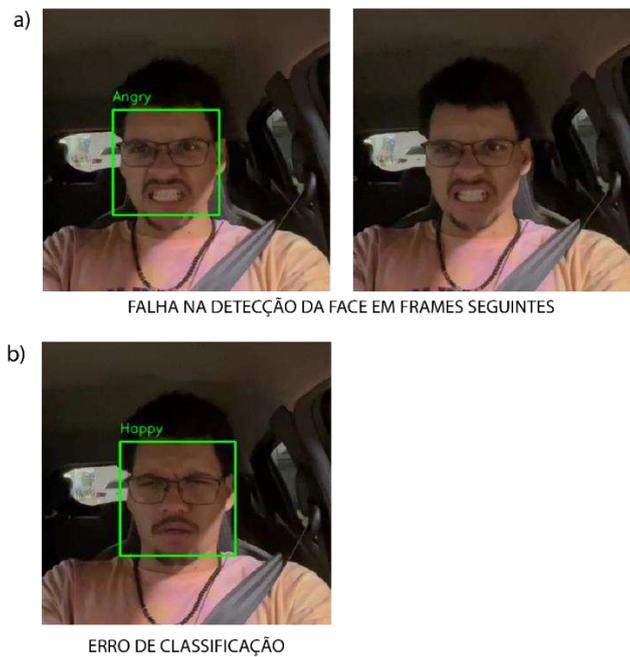
Figura 14 – Dificuldade na detecção do rosto por causa dos óculos.



Fonte: Elaborado pelo autor.

Os problemas com iluminação se destacam ainda mais em ambientes com baixa luminosidade, como foi possível notar nos testes feitos durante a noite. Neles, ocorreu a incidência de diversos *frames* sem a detecção do rosto, como visto na imagem “a” da Figura 15 e a classificação de emoções erradas com maior frequência, presente na imagem “b” da Figura 15.

Figura 15 – Erros recorrentes nos testes feitos no período noturno com óculos.



Fonte: Elaborado pelo autor.

Nos testes feitos sem óculos no mesmo período, obteve-se os mesmos problemas, conforme a Figura 16.

Figura 16 – Erros recorrentes nos testes feitos no período noturno sem óculos.



Fonte: Elaborado pelo autor.

## 6 CONCLUSÃO

Os elevados índices de acidentes de trânsito no Brasil e no mundo demonstram a importância do desenvolvimento de pesquisas, com o objetivo de produzir recursos que contribuam com a segurança de todos neste meio.

Com isso, o uso de ferramentas computacionais neste cenário tem evoluído em técnicas cada vez mais funcionais. O resultado deste trabalho é um modelo de rede neural, que tem como intuito identificar as emoções do condutor e fazê-las aplicáveis, tornando-as presente no cotidiano.

Este trabalho desenvolveu uma metodologia para a identificação de emoções em imagens através das expressões faciais. Após isso, foi feita uma análise em 4 cenários diferentes, onde o motorista dirigia com e sem óculos e nos períodos de dia e de noite, desta forma, expondo o comportamento da RNA.

Os resultados apresentados no Capítulo 4 demonstram o desempenho promissor da metodologia desenvolvida na tarefa de classificação de emoções. Foram obtidos valores de 82,82% de precisão média, 80,87% de sensibilidade média e 81,00% de acurácia média em todas as 7 emoções.

### 6.1 Trabalhos Futuros

É importante destacar que uma das dificuldades do trabalho foi em relação à própria base de dados utilizada, pelo fato que a grande maioria das imagens possuem expressões faciais muito caricatas. Ademais, o identificador de faces utilizado não se demonstrou tão eficaz nos testes feitos durante a noite e com óculos. No entanto, ainda assim, existem fatores que indicam a necessidade da ampliação e desenvolvimento deste trabalho. Nesse sentido são colocados alguns aspectos de melhoria:

- Criar ou usar um identificador de faces que seja eficaz nos cenários onde as pessoas usam acessórios, tais como chapéu e óculos.
- Utilização de outras expressões e emoções capazes de comprometer a segurança no trânsito, como ansiedade, estresse, sonolência e cansaço.
- Utilização de transferência de aprendizagem para melhores resultados.
- Utilização do Local Binary Patterns (LBP), que é útil para o

reconhecimento de padrões, assim fazendo a extração de características de texturas em imagens.

Por fim, a premissa apresentada neste trabalho poderá ser integrada a outros softwares e sensores presentes nos veículos. Dessa forma, será possível aprimorar a eficiência e a segurança das operações, oferecendo uma experiência de condução mais avançada e automatizada. A integração com sistemas de assistência ao motorista, como controle de cruzeiro adaptativo e frenagem de emergência, permitirá uma maior interação entre os diferentes componentes do veículo, resultando em um melhor desempenho e na redução de acidentes. A combinação dessas tecnologias promoverá uma condução mais inteligente e eficiente, beneficiando tanto os motoristas quanto a sociedade como um todo.

## REFERÊNCIAS

ACAYABA, Edvaldo et al. Carros Inteligentes.

ADOLPHS, Ralph. Recognizing emotion from facial expressions: psychological and neurological mechanisms. **Behavioral and cognitive neuroscience reviews**, v. 1, n. 1, p. 21-62, 2002.

ADOLPHS, Ralph; TRANEL, Daniel; DAMASIO, Antonio R. Dissociable neural systems for recognizing emotions. **Brain and cognition**, v. 52, n. 1, p. 61-69, 2003.

BASKIN-SOMMERS, Arielle; STUPPY-SULLIVAN, Allison M.; BUCKHOLTZ, Joshua W. Psychopathic individuals exhibit but do not avoid regret during counterfactual decision making. **Proceedings of the National Academy of Sciences**, v. 113, n.50, p. 14438-14443, 2016.

BEHLING, Alexandre et al. Reconhecimento de emoções em vídeo utilizando redes neurais artificiais. 2019.

BEZERRA, Eduardo. Introdução à aprendizagem profunda. **Artigo-31º Simpósio Brasileiro de Banco de Dados-SBBD2016-Salvador**, 2016.

COSTA, Jôberth Rogers Tavares. Aplicação de Deep Learning para análise de sentimentos a partir de imagens. 2022.

DAMACENO, Siuari Santos et al. Inteligência artificial: uma breve abordagem sobre seu conceito real e o conhecimento popular. **Caderno de Graduação-Ciências Exatas e Tecnológicas-UNIT-SERGIPE**, v. 5, n. 1, p. 11-11, 2018

DARAPANENI, Narayana et al. Detection of Distracted Driver using Convolution Neural Network. **arXiv preprint arXiv:2204.03371**, 2022.

DE ÁVILA NEGRI, Sergio Marcos Carvalho; DE OLIVEIRA, Samuel Rodrigues; COSTA, Ramon Silva. O uso de tecnologias de reconhecimento facial baseadas em inteligência artificial e o direito à proteção de dados. **Direito Público**, v. 17, n. 93, 2020.

DE OLIVEIRA, Eduardo; JAQUES, Patrícia Augustin. Classificação de emoções básicas através de imagens capturadas por webcam. **Revista Brasileira de Computação Aplicada**, v. 5, n. 2, p. 40-54, 2013.

EKMAN, Paul. **A Linguagem das emoções**. São Paulo: Lua de Papel, 2011.

**Emotional Recognition**. Murilo Marin Pechoto, 2017. Disponível: <https://github.com/mpechoto/EmotionRecognition>. Acesso em: 14 mar. 2023.

FUNÇÃO de ativação. Deep Learning. Disponível em:

<https://www.deeplearningbook.com.br/funcao-de-ativacao/>. Acesso em: 28 mai. 2023.

GONZALEZ, Rafael C.; WOODS, Richard E. **Processamento de imagens digitais**. Editora Blucher, 2000.

GOODFELLOW, Ian J. et al. Challenges in representation learning: A report on three machine learning contests. In: **Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013. Proceedings, Part III 20**. Springer berlin heidelberg, 2013. p. 117-124.

HEATON, Jeff. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618. **Genetic Programming and Evolvable Machines**, v. 19, n. 1-2, p. 305-307, 2018.

HUANG, Weijie; WANG, Jun. Character-level convolutional network for text classification applied to chinese corpus. **arXiv preprint arXiv:1611.04358**, 2016.

JAIN, Anil K.; LI, Stan Z. **Handbook of face recognition**. New York: springer, 2011.

JANIESCH, Christian; ZSCHECH, Patrick; HEINRICH, Kai. Machine learning and deep learning. **Electronic Markets**, v. 31, n. 3, p. 685-695, 2021.

KANADE, Takeo; COHN, Jeffrey F.; TIAN, Yingli. Comprehensive database for facial expression analysis. In: **Proceedings fourth IEEE international conference on automatic face and gesture recognition (cat. No. PR00580)**. IEEE, 2000. p. 46-53.

LAUANDE, Marcos Gabriel Mendes et al. Classificação de imagens histopatológicas de câncer de pênis usando redes convolucionais e transferência de aprendizagem. 2022.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **nature**, v. 521, n. 7553, p. 436-444, 2015.

MATOS, Hamilton de et al. Desenvolvimento e validação de um sistema de identificação de emoções por visão computacional e redes neurais convolucionais com transferência de aprendizado. 2017. 50 f. Dissertação (Engenharia Elétrica) - Universidade Presbiteriana Mackenzie, São Paulo.

LOPES, Eduardo Costa; BINS FILHO, José Carlos; NO, RELATÓRIO TÉCNICO. Detecção de faces e características faciais. **Porto Alegre: PUCRS**, 2005.

MONARD, Maria Carolina e BARANAUSKAS, José Augusto. Aplicações de inteligência artificial: uma visão geral. **Anais**, 2000.

PASSOS, Bianca Tallita. O mundo do ponto de vista das Redes Neurais Convolucionais (RNCs). **Blog Ateliware**, 2021. Disponível em: <https://ateliware.com/blog/redes-neurais-convolucionais>. Acesso em: 08 jun. 2023.

PISAREVSKY, Vadim. Haar Cascade Frontal Face Default. GitHub, 2013. Disponível

em:

[https://github.com/opencv/opencv/blob/master/data/haarcascades/haarcascade\\_frontalface\\_default.xml](https://github.com/opencv/opencv/blob/master/data/haarcascades/haarcascade_frontalface_default.xml). Acesso em: 14 mar. 2023.

SCARDUA<sup>1</sup>, Demétrius Alex; MARQUES, Kédyma. **Estudo da Identificação de Emoções Através da Inteligência Artificial**. 2018.

SCHROFF, Florian; KALENICHENKO, Dmitry; PHILBIN, James. Facenet: A unified embedding for face recognition and clustering. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2015. p. 815-823.

SHORTEN, Connor; KHOSHGOFTAAR, Taghi M. A survey on image data augmentation for deep learning. **Journal of big data**, v. 6, n. 1, p. 1-48, 2019.

SONG, Jia et al. A survey of remote sensing image classification based on CNNs. **Big earth data**, v. 3, n. 3, p. 232-254, 2019.

SZELISKI, Richard. **Computer vision: algorithms and applications**. Springer Nature, 2022.

VISUALIZAÇÃO da matriz de confusão. **IBM**, 2021. Acesso em: <https://www.ibm.com/docs/pt-br/db2/10.5?topic=visualizer-confusion-matrix-view>. Acesso em: 28 mai. 2023.

YE, Lanhang; YAMAMOTO, Toshiyuki. Evaluating the impact of connected and autonomous vehicles on traffic safety. **Physica A: Statistical Mechanics and its Applications**, v. 526, p. 121009, 2019.

ZEPF, Sebastian et al. Driver emotion recognition for intelligent vehicles: A survey. **ACM Computing Surveys (CSUR)**, v. 53, n. 3, p. 1-30, 2020.

ZHANG, Aston et al. Dive into deep learning. **arXiv preprint arXiv:2106.11342**, 2021.