

Aprimorando interações humano-computador em aplicações com reconhecimento de emoções por redes neurais convolucionais

Maria Eduarda Lima Santos*

**Universidade Federal do Recôncavo da Bahia, Centro de Ciências Exatas e Tecnológicas,
Cruz das Almas, Bahia, Brasil (E-mail: marialima@aluno.ufrb.edu.br)*

Abstract: This article explores the use of convolutional neural networks for the recognition of emotions from facial expressions, focusing on human-computer interaction. Emotion recognition is essential for personalizing and enhancing the user experience. The implementation of AI algorithms for identifying emotions through facial expressions was investigated, and the technical and commercial viability of implementing this tool in various applications was evaluated. This study aims to enhance user satisfaction, paving the way for innovations in various areas of human life.

Resumo: Este artigo explora o uso de redes neurais convolucionais para o reconhecimento de emoções a partir de expressões faciais, com foco na interação humano-computador. O reconhecimento de emoções é fundamental para personalizar e aprimorar a experiência do usuário. Investigou-se a implementação de algoritmos de IA para identificar emoções por meio de expressões faciais e avaliou-se a viabilidade técnica e comercial da implementação desta ferramenta em diversas aplicações. Este estudo visa aprimorar a satisfação do usuário, abrindo caminhos para inovações em diversas áreas da vida humana.

Keywords: Convolutional neural networks; Emotion recognition; Facial expressions; Human-computer interaction; AI algorithms; User experience.

Palavras-chaves: Redes neurais convolucionais; Reconhecimento de emoção; Expressões faciais; Interação humano-computador; Algoritmos de IA; Experiência de usuário

1. INTRODUÇÃO

Nos últimos anos, estamos testemunhando uma notável revolução na forma como a tecnologia e a inteligência artificial (IA) estão remodelando nossas interações com sistemas computacionais. Esse cenário emocionante e promissor no campo da IA é especialmente evidente no reconhecimento de emoções a partir de expressões faciais. A capacidade das máquinas em compreender as emoções humanas abre novas perspectivas para uma interação humano-computador (IHC) mais natural, adaptativa e eficaz. Atualmente, com a integração cada vez mais profunda de computadores e dispositivos eletrônicos em nossas vidas, a ênfase na criação de interfaces "amigáveis" para a interação humano-computador tem crescido significativamente (CYBIS; BETION; FAUST, 2010). A interface atua como o ponto de conexão entre os usuários e a tecnologia, desempenhando um papel central em qualquer sistema computacional. Essa interação pode ocorrer de várias maneiras, incluindo interações físicas, perceptivas e conceituais (MORAN, 1981).

Nesse contexto, a compreensão das emoções humanas desempenha um papel crucial na otimização da experiência do usuário. Isso permite que as interfaces se adaptem de forma mais eficaz às necessidades e estados emocionais dos

usuários. Portanto, o reconhecimento de emoções a partir de expressões faciais se destaca como um elemento fundamental, fornecendo uma base sólida para o desenvolvimento de sistemas de IHC mais intuitivos e sensíveis às emoções humanas.

1.1 Trabalhos relacionados

Melo et al. conduziram um estudo comparativo de técnicas computacionais para a classificação de emoções, explorando métodos para detectar automaticamente seis emoções básicas em expressões aparentes usando o banco de dados Cohn-Kanade. As técnicas comparadas incluíram Redes Bayesianas, Máquinas de Vetores de Suporte e Árvores de Decisão, com resultados de alta precisão para Redes Bayesianas. Este estudo destacou aplicações potenciais em psicologia e inteligência artificial, melhorando o reconhecimento e compreensão de emoções.

Por outro lado, Oliveira et al. apresentaram um sistema computacional que identifica emoções básicas (raiva, nojo, medo, surpresa, alegria e tristeza) por meio de expressões faciais do usuário. O processo detalhado de remoção de dados de características e a identificação de pontos nessas características foram planejados. O sistema foi testado em vídeos de baixa resolução capturados por uma webcam,

demonstrando a capacidade de identificar emoções de forma não intrusiva.

1.2 Justificativa

Este estudo fundamenta a necessidade de aprimorar a experiência do usuário, tornando as interações mais personalizadas, precisas e envolventes. Através do reconhecimento de emoções a partir de expressões faciais, buscamos aprimorar a interação humano-computador, tornando-a mais natural e intuitiva, permitindo que os sistemas compreendam e se adaptem às emoções dos usuários. Essa abordagem tem vastas aplicações multidisciplinares, incluindo setores críticos como saúde, onde o monitoramento emocional pode beneficiar pacientes, educação, possibilitando a personalização do ensino, entretenimento, adaptando experiências de jogos e muito mais.

1.3 Objetivos

- Explorar como as redes neurais convolucionais podem ser implementadas em sistemas de interação humano-computador para reconhecer e interpretar as emoções dos usuários em tempo real.
- Explorar como as redes neurais convolucionais podem ser implementadas em sistemas de interação humano-computador para reconhecer e interpretar as emoções dos usuários em tempo real.
- Implementar algoritmos de IA que possam ser capazes de identificar e classificar as emoções através das expressões faciais dos usuários.
- Realizar uma avaliação teórica técnica e comercial para compreender o impacto da aplicação desse modelo.

2. FUNDAMENTAÇÃO TEÓRICA

2.1.1 Emoção e expressões faciais

As emoções humanas são respostas automáticas, universais e adaptativas expressas por meio de expressões faciais e outras manifestações comportamentais. São inatas e desencadeiam respostas comportamentais específicas, semelhantes em várias culturas (EKMAN, 2011).

A maneira como as emoções são demonstradas pode ser expressa de formas variadas, tais como linguagem corporal, tom de voz, velocidade e ritmo da fala, além das expressões faciais que são definidas como padrões de movimentos musculares específicos que ocorrem na face e estão associados a emoções, estados emocionais ou intenções comunicativas (EKMAN, 2011).

Uma das classificações de emoções mais influentes é a de Paul Ekman, com seis emoções básicas universais definidas como: raiva, nojo/aversão, medo, felicidade, tristeza, surpresa. Além delas, existe também a emoção neutra. Essas emoções desempenham um papel crucial na experiência humana e são consideradas universais porque são encontradas e reconhecidas em diferentes culturas ao redor do mundo (EKMAN, 2011).

2.2 Visão computacional

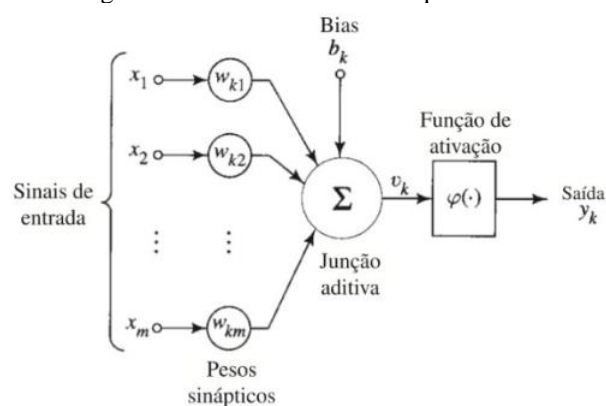
A visão computacional é um campo de inteligência artificial que se concentra no desenvolvimento de algoritmos e técnicas que permitem aos computadores "verem" e compreenderem o mundo visual de maneira semelhante aos seres humanos (Russel e Norvig, 1995). Segundo Marr (apud. Russel e Norvig), a visão é o processo de descobrir, por meio de imagens, o que está presente no mundo e onde está. A visão computacional busca auxiliar na resolução de problemas complexos, simulando a visão humana, capturando, processando e analisando imagens, imitando a habilidade do ser humano de tomar decisões com base em informações visuais. Ela é responsável por permitir que o computador interprete o ambiente ao seu redor, monitorando e manipulando objetos nas imagens capturadas. Essa tecnologia encontra aplicações em diversos setores, como agronomia, biologia, medicina, indústria, segurança e educação (NEVES; NETO; GONZAGA, 2012).

2.3 Redes Neurais Artificiais (RNA's)

2.3.1 Inspiração Biológica e Neurônio Artificial

O estudo das redes neurais surgiu da compreensão de que o cérebro humano opera de forma distinta das computadores digitais tradicionais (HAYKIN, 2007). O cérebro é um sistema complexo, não linear e paralelo, composto por neurônios que realizam processamentos específicos. Essas unidades respondem a estímulos por meio de mudanças na diferença de potencial elétrico em suas membranas celulares, transmitindo essas respostas para outras células (FURTADO, 2019). O neurônio artificial, que são a unidade de processamento em redes neurais, replicam a estrutura e o funcionamento dos neurônios biológicos, com dendritos como entradas, o soma para processamento e o axônio como saída (FURTADO, 2019).

Figura 1 - Neurônio artificial esquematizado



Fonte: HAYKIN, 2007, p.36

2.3.3 Definição de Rede Neural Artificial

Uma Rede Neural Artificial (RNA) é um sistema conector composto por neurônios interconectados, processando informações de maneira distribuída e simultânea

(FURTADO, 2019). Essa estrutura permite um processamento colaborativo e a identificação de padrões em conjuntos de informações. As RNA podem ser rompidas eletronicamente ou por programação em computadores digitais. Elas acumularam conhecimento empírico, tornando-o acessível para uso (HAYKIN, 2007).

2.3.4 Funções de Ativação

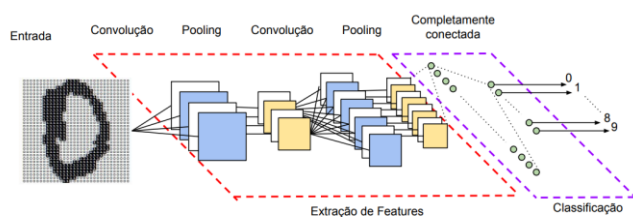
As funções de ativação moldam o valor de saída dos neurônios em redes neurais (FURTADO, 2019). Algumas funções relevantes para este trabalho incluem:

- Função Limiar: Ativa um neurônio se a entrada ultrapassa um limite (HAYKIN, 2007).
- Unidade Linear Retificada (ReLU): Retorna a entrada se for positiva, caso contrário, zero (MAAS; HANNUN; NG, 2013).
- Função Softmax: Usada em problemas de classificação, mapeia um vetor para probabilidades (WANG et al., 2018).

2.3.5 Treinamento e Aprendizagem

O treinamento de uma Rede Neural Artificial envolve ajustar suas configurações por meio de algoritmos de treinamento (FURTADO, 2019). Existem três categorias de aprendizagem: supervisionado, não supervisionado e por reforço (FURTADO, 2019). O aprendizado supervisionado envolve treinar a rede com pares de dados de entrada e saída conhecidos, ajustando os parâmetros até minimizar o erro. O aprendizado não supervisionado analisa dados para identificar semelhanças e padrões. O aprendizado por reforço otimiza a tomada de decisões sem respostas corretas predefinidas, baseando-se em feedback externo (FURTADO, 2019).

Figura 2 – Exemplo de CNN



Fonte: VARGAS et al., 2016, p.1

2.4 Redes Neurais Convolucionais (CNN's)

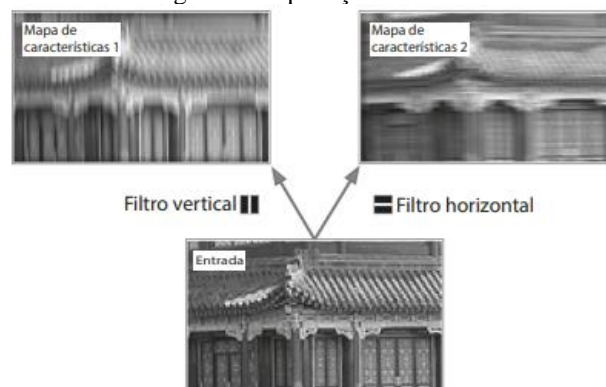
As Redes Neurais Convolucionais (CNNs), mostrada na Fig. 2, revolucionaram o processamento de imagens. Inspiradas no cérebro humano, essas redes imitam como interpretamos informações visuais. Com base na estrutura do córtex visual, as CNNs podem importar objetos, identificar rostos, expressões superficiais e cenas (GÉRON, 2019).

2.4.1 Filtros e operação de convolução

Os filtros na CNN, como mostrado da Fig. 3, são conjuntos de pesos usados na operação de convolução para destacar

padrões nas imagens. Eles são fundamentais para identificar características importantes nas imagens (GÉRON, 2019).

Figura 3 – Aplicação de filtros



Fonte: GÉRON, 2019, p.374

2.4.2 Mapas de características

A operação de convolução cria mapas de características, que representam como características identificadas pelos filtros estão distribuídos na imagem. Isso é crucial para a identificação de padrões em imagens (PATTANAYAK, 2017).

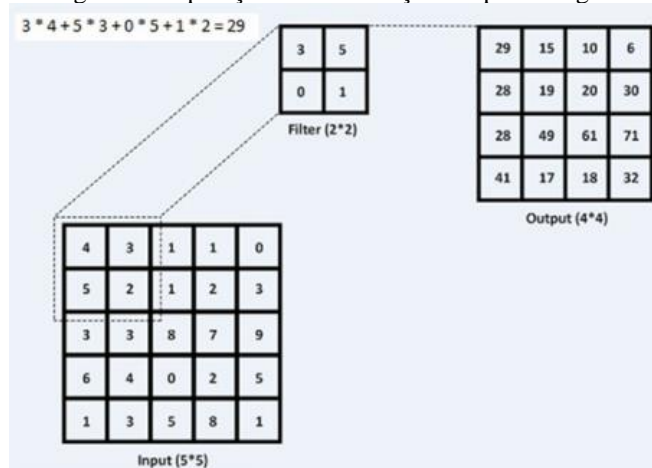
2.4.3 Camada de convolução

A camada de convolução aplica operações de convolução, como mostrado na Fig. 4, usando filtros para identificar padrões nas imagens. Os filtros têm pesos ajustados durante o treinamento da rede (PATTANAYAK, 2017).

2.4.4 Camada de pooling

A camada de pooling reduz a dimensionalidade das representações de características, preservando informações relevantes e simplificando o processamento (PATTANAYAK, 2017).

Figura 4 – Operação de Convolução 2D para imagem



Fonte: MANASWI, 2018, p. 93

2.4.5 Camadas totalmente conectadas

Camadas totalmente conectadas realizando a classificação final com base nas características extraídas. Cada neurônio está conectado a todos os neurônios na camada anterior, permitindo a captura de relações complexas (PATTANAYAK, 2017).

2.5 IHC

Campo de estudo que se concentra na interação entre seres humanos e sistemas computacionais, para tornar essa interação mais eficaz, eficiente e eficiente. (BARBOSA; SILVA, 2010).

3. METODOLOGIA

Este trabalho se baseia na implementação de redes neurais convolucionais para o reconhecimento de emoções por expressões faciais, voltada ao aprimoramento da interação humano-computador em aplicações.

3.1 Ferramentas e instrumentos

Google Drive: plataforma de armazenamento em nuvem desenvolvida pelo Google, frequentemente usada para armazenar e compartilhar grandes conjuntos de dados e modelos de aprendizado de máquina.

Google Colaboratory (Colab): plataforma de desenvolvimento baseada em nuvem que oferece recursos de GPU e TPU para acelerar o treinamento de modelos de aprendizado profundo.

VsCode (Visual Studio Code): editor de código amplamente utilizado desenvolvido pela Microsoft. É uma ferramenta de código aberto que oferece suporte a diversas linguagens de programação.

Banco de dados FER2013: O conjunto de dados FER2013 consiste em 35.887 imagens em tons de cinza, cada uma com resolução de 48x48 pixels e categorizadas em sete emoções distintas: raiva, nojo, medo, felicidade, neutro, tristeza e surpresa. Este banco de dados surgiu a partir de uma competição realizada na plataforma Kaggle em 2013 e se tornou amplamente utilizado como referência para treinamento, validação e teste de modelos de aprendizado de máquina voltados ao reconhecimento de emoções. Ele é fornecido em formato CSV, onde as colunas atribuem códigos numéricos de 0 a 6 para representar as diferentes emoções e sequências de números que indicam os valores de escala de cinza para cada pixel nas imagens.

Bibliotecas e API do Python (TensorFlow, OpenCV, OS, NumPy, Keras, matplotlib.pyplot): Um conjunto de bibliotecas que desempenham papéis essenciais no desenvolvimento de projetos de ciência de dados e aprendizado de máquina. Isso inclui TensorFlow para aprendizado de máquina e aprendizado profundo, OpenCV para visão computacional, OS para interagir com o sistema

operacional, NumPy para computação numérica, Keras para construção e treinamento de redes neurais profundas e matplotlib.pyplot para visualização de dados, permitindo a criação de gráficos e plots informativos.

3.2 Abordagem da Pesquisa

A abordagem da pesquisa utilizada neste trabalho é predominantemente quantitativa, pois se concentra na análise das imagens do conjunto de dados FER2013, que são representadas numericamente como matrizes de valores de tons de cinza dos pixels e na aplicação de algoritmos de aprendizado de máquina.

3.3 População e Amostra

A população de interesse é composta por 35.887 imagens do conjunto de dados FER2013 categorizadas em sete emoções representadas por números de 0 a 6 (0-raiva, 1-nojo, 2-medo, 3-felicidade, 4-tristeza, 5-surpresa, 6-neutro), apresentada na Fig 5.

Figura 5 – Banco de dados do FER2013

	emotion	pixels	Usage
35882	6	50 36 17 22 23 29 33 39 34 37 37 39 43 48 5...	PrivateTest
35883	3	178 174 172 173 181 188 191 194 196 199 200 20...	PrivateTest
35884	0	17 17 16 23 28 22 19 17 25 26 20 24 31 19 27 9...	PrivateTest
35885	3	30 28 28 29 31 30 42 68 79 81 77 67 67 71 63 6...	PrivateTest
35886	2	19 13 14 12 13 16 21 33 50 57 71 84 97 108 122...	PrivateTest

Fonte: Imagem obtida em análise feita pelo GoogleColab

3.4 Cronograma

Fase 1: Exploração das Bases Teóricas Relevantes para o Projeto.

Fase 2: Pesquisa de estudos anteriores relacionados ao tema.

Fase 3: Escolha e Aquisição dos Conjuntos de Dados Necessários.

Fase 4: Organização e Formatação dos Dados para Análise; Criação de Algoritmos para Treinamento da Rede Neural; Implementação das técnicas e métodos de treinamento.

Fase 5: Realização do Processo de Treinamento com os Dados.

Fase 6: Avaliação do Desempenho do Modelo e Interpretação dos Resultados; Análise das métricas de desempenho; Síntese das Descobertas e Insights Obtidos

3.5 Procedimentos:

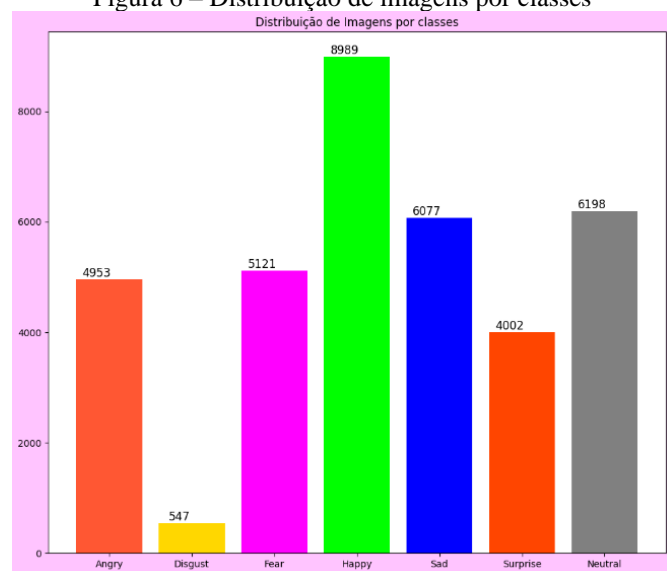
Pesquisa Teórica: inicialmente, fez-se uma pesquisa teórica abrangente para compreender os conceitos fundamentais relacionados ao tema. Além disso, investigou-se trabalhos anteriores relevantes que foram desenvolvidos para o embasamento teórico do projeto.

Aquisição do Banco de Dados: O conjunto de dados foi obtido no Kaggle e, posteriormente, transferido para o Google Drive, onde se tornou acessível para análise e uso nos experimentos. Essa etapa é crucial para garantir a disponibilidade dos dados essenciais.

Bibliotecas Importadas: No ambiente do Colab, importam-se diversas bibliotecas fundamentais, sendo o OpenCV (cv2) essencial para o processamento de imagens, o NumPy utilizado para manipulação eficiente de arrays e dados numéricos, o Pandas para facilitar a operação de dados em formato CSV, o Matplotlib como uma ferramenta crucial para visualização de informações, e o TensorFlow desempenha um papel central na construção e treinamento dos modelos. Além disso, outras bibliotecas relacionadas também são importadas para funções específicas dentro do projeto.

Leitura, análise e pré-processamento do Conjunto de Dados : O arquivo CSV do conjunto de dados “fer2013.csv”, é lido e a base é acessada usando a biblioteca Pandas. Esse conjunto de dados contém informações sobre expressões faciais rotuladas e é utilizado para treinar e avaliar o modelo de reconhecimento, como apresentado na Fig 5. Realiza-se uma análise da distribuição das expressões emocionais no conjunto de dados, apresentada na Fig. 6 pelo gráfico de barras, que contém quantas amostras de cada emoção são apresentados no conjunto de dados e exibem o resultado. Logo após, converte-se as sequências de pixels em arrays NumPy, normaliza-se os valores dos pixels para o intervalo [0, 1] e expande as dimensões das imagens para que possam ser usadas como entrada para o modelo.

Figura 6 – Distribuição de imagens por classes

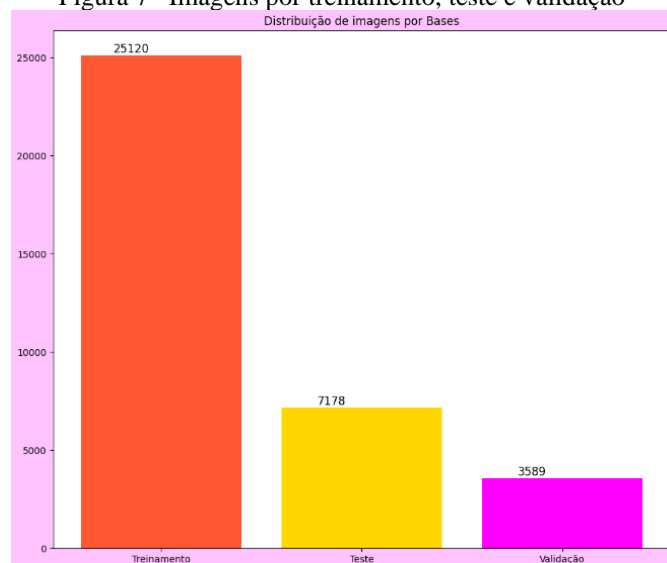


Fonte: Imagem obtida em análise feita pelo GoogleColab

Divisão do Conjunto de Dados: O conjunto é dividido em três partes - treinamento (70%), teste (20%), e validação (10%). Isso é realizado com a função `train_test_split` do Scikit-Learn, sendo fundamental para construir, avaliar e aprimorar modelos de aprendizado de máquina e a amostragem de cada parte é apresentado na Fig 7.

Construção, Treinamento e Avaliação do Modelo: O modelo de reconhecimento de expressões visuais é construído com TensorFlow e Keras, usando uma rede neural convolucional (CNN) para extrair características das imagens. O treinamento ocorre ao longo de 100 épocas, usando o otimizador Adam e a função de perda de entropia cruzada categórica. Durante o treinamento, são aplicados callbacks, como `ReduceLROnPlateau`, `EarlyStopping` e `ModelCheckpoint`, para ajustar a taxa de aprendizado, evitar overfitting e salvar o modelo com melhor desempenho na validação. Gráficos são gerados para acompanhar a acurácia e perda ao longo do treinamento. Após o treinamento, o modelo é avaliado e os resultados são visualizados.

Figura 7– Imagens por treinamento, teste e validação



Fonte: Imagem obtida em análise feita pelo GoogleColab

4. RESULTADOS E DISCUSSÕES

Neste estudo, conduzimos uma avaliação minuciosa do desempenho de uma Rede Neural Convolucional (CNN) no contexto do reconhecimento de emoções.

Os resultados da base de treinamento e validação são apresentados em detalhes nos gráficos de Acurácia e Perda (Loss) na Figura 8. Estes gráficos contêm informações essenciais para a avaliação do desempenho da CNN no reconhecimento emocional.

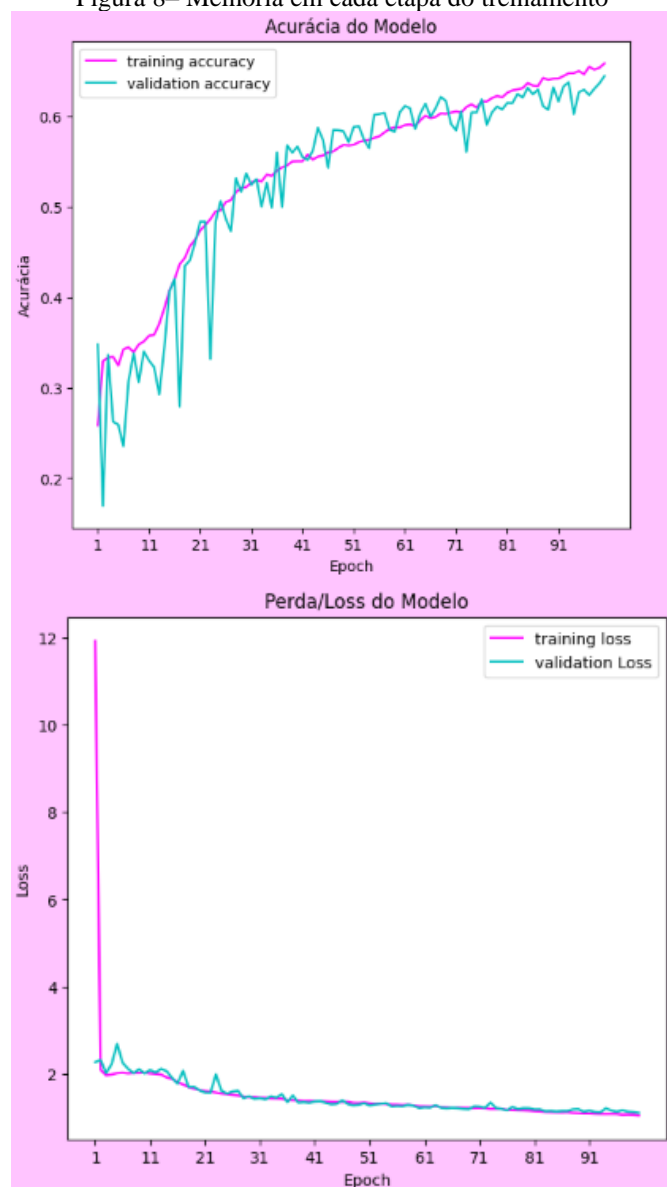
A curva de Acurácia quantifica a capacidade do modelo de classificação de emoções, enquanto os gráficos de Perda (Loss) indicam como o erro do modelo diminui ao longo do treinamento. Um destaque do estudo é o resultado da Epoch 100, que revela uma melhoria na `val_loss`, passando de 1.11676 para 1.10613, momento em que o modelo foi salvo como `modelo_expressoes1.h5`. Essa métrica ilustra a evolução do modelo ao longo do treinamento, apresentando uma perda de 1,0456 e uma acurácia de 65,84% no conjunto de treinamento, bem como uma `val_loss` de 1,1061 e uma `val_accuracy` de 64,45% no conjunto de validação. A taxa de aprendizagem adotada foi de 2.2877e-04. Essas análises são de importância crítica para a compreensão do desempenho do modelo no reconhecimento de emoções.

Os resultados obtidos a partir da base de teste revelaram uma acurácia de aproximadamente 62,48%, o que indica a capacidade da CNN em avaliar corretamente as emoções em imagens visíveis. A perda (loss) associada a essas variações foi registrada em 1,1361.

É importante notar que a acurácia de 62,12% é um indicativo positivo, considerando que o reconhecimento de emoções em imagens é uma tarefa necessária devido à complexidade das expressões ósseas humanas. No entanto, a análise do desempenho do modelo não deve ser limitada à acurácia.

A matriz de confusão, apresentada na Fig 9, revelou que a rede foi capaz de prever corretamente mais da metade dos dados de teste, evidenciando sua capacidade de identificar com sucesso diversas emoções.

Figura 8– Melhoria em cada etapa do treinamento



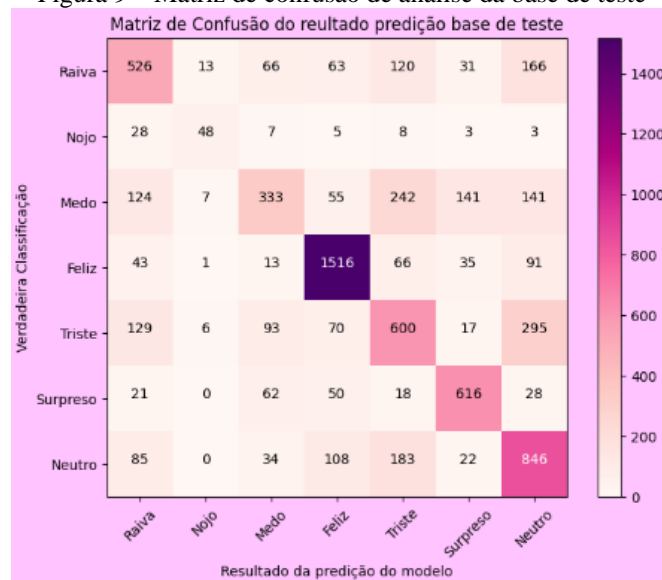
Fonte: Imagem obtida em análise feita pelo GoogleColab

A avaliação da matriz de confusão é fundamental para compreender como o modelo se comporta em relação a cada classe emocional. Identificar quais emoções são mais

facilmente reconhecidas e quais apresentam maiores desafios pode orientar melhorias no modelo.

No estudo identificou-se poucas amostras da emoção nojo, no banco, em comparação as demais emoções, entretanto, a quantidade da amostra não impactou tão significativamente a capacidade do modelo em reconhecer a emoção em questão.

Figura 9 – Matriz de confusão de análise da base de teste



Fonte: Imagem obtida em análise feita pelo GoogleColab

Os resultados indicam que a aplicação de um modelo de Rede Neural Convolutiva (CNN) para o reconhecimento emocional em várias aplicações é promissora tanto do ponto de vista técnico quanto comercial. A liberdade da CNN permite sua integração em uma ampla gama de plataformas, desde aplicativos móveis e serviços de streaming até dispositivos de realidade virtual. Ela é capaz de processar grandes volumes de dados, identificar padrões complexos em imagens e fornecer classificações de alta precisão em tempo real.

A adaptabilidade da CNN não se limita à sua aplicação, uma vez que pode ser treinada para tratar diversas expressões emocionais, tornando-se altamente personalizável para atender a diferentes necessidades. Do ponto de vista comercial, essa tecnologia tem potencial em segmentos como entretenimento, publicidade e saúde mental.

Nas plataformas de streaming, o reconhecimento emocional pode aumentar a relevância do conteúdo, impulsionando o consumo e a receita. A publicidade direcionada com base nas emoções dos usuários também se torna mais eficaz. Além disso, aplicativos de saúde mental podem oferecer suporte personalizado com base no estado emocional do usuário, criando oportunidades de negócios para desenvolvedores de aplicativos.

A Internet das Coisas (IoT) também pode se beneficiar ao entender as emoções dos usuários, permitindo configurações personalizadas, como termostatos inteligentes adaptando a temperatura de acordo com o humor do usuário. No entanto, é importante abordar considerações éticas e de privacidade relacionadas à coleta e uso de dados emocionais para garantir

a confiança do público. A aplicação do reconhecimento emocional por meio da CNN mostra-se promissora tanto do ponto de vista técnico quanto comercial, aprimorando experiências do usuário e criando oportunidades no mercado.

5. CONCLUSÃO

Neste estudo, avaliou-se o desempenho de uma Rede Neural Convolucional (CNN) no reconhecimento de emoções. Os gráficos de Acurácia e Perda mostraram que a CNN desenvolvida apresentou uma perda de 1,0456 e uma acurácia de 65,84% no conjunto de treinamento, bem como uma val_loss de 1,1061 e uma val_accuracy de 64,45% no conjunto de validação e obteve uma pontuação de cerca de 62,12% no teste e uma perda de 1,2097. A análise da matriz de confusão revelou que a CNN foi capaz de prever corretamente mais da metade dos dados de teste, demonstrando sua capacidade de identificar diferentes emoções.

A aplicação da CNN no reconhecimento emocional em várias áreas é promissora tanto do ponto de vista técnico quanto comercial. Ela é altamente adaptável, podendo ser integrada em uma variedade de plataformas e oferece processamento eficiente de dados e classificações em tempo real. Além disso, pode ser treinada para refletir sobre várias expressões emocionais, tornando-a personalizável para atender a diferentes necessidades. Do ponto de vista comercial, essa tecnologia tem potencial em setores como entretenimento, publicidade e saúde mental, melhorando a relevância do conteúdo, a eficácia da publicidade e o suporte personalizado. A Internet das Coisas (IoT) também pode se beneficiar com a compreensão das emoções dos usuários.

No entanto, questões éticas e de privacidade relacionadas à coleta e uso de dados emocionais devem ser consideradas. Em resumo, o reconhecimento emocional por meio da CNN oferece oportunidades especiais para aprimorar a experiência do usuário e inovações em diversas áreas.

REFERÊNCIAS

- EKMAN, P. (2011). A linguagem das emoções. Lua de papel. ISBN-10: 8563066420, ISBN-13: 978-8563066428.
- BARBOSA, Simone Diniz Junqueira; SILVA, Bruno Santana da. Interação humano-computador. Rio de Janeiro: Elsevier Editora Ltda, 2010. Recurso digital. ISBN 978-85-352-1120-7.
- CYBIS, W.; BETION, A. H.; FAUST, R. Ergonomia e Usabilidade: conhecimentos, métodos e técnicas. [S.l.]: Novatec Editora, 2010.
- DE MELO, Sara L. et al. Estudo comparativo de técnicas computacionais para classificação de emoções. In: Simpósio Brasileiro de Informática na Educação (Simpósio Brasileiro de Informática na Educação-SBIE) . 2014. pág. 456
- 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
- FURTADO, M. I. V. Redes neurais artificiais: uma abordagem para sala de aula. Ponta Grossa, PR: Atena Editora, 2019
- GÉRON, Aurélien. Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow: Conceitos, Ferramentas e Técnicas para a Construção de Sistemas Inteligentes . Traduzido por Rafael Contatori. Rio de Janeiro: Alta Livros, 2019. 576 p. ISBN: 978-85-508-0902-1 (PDF).
- HAYKIN, S. Redes Neurais: Princípios e Prática. [S.l.]: Artmed, 2007. ISBN 9788577800865
- MAAS, A. L.; HANNUN, A. Y.; NG, A. Y. Rectifier nonlinearities improve neural network acoustic models. In: in ICML Workshop on Deep Learning for Audio, Speech and Language Processing. [S.l.: s.n.], 2013
- MANASWI, Navin Kumar. Aprendizado profundo com aplicativos usando Python: chatbots e reconhecimento facial, de objetos e de fala com TensorFlow e Keras. Bangalore, Karnataka, Índia, 2018. ISBN-13 (pbk): 978-1-4842-3515-7. ISBN-13 (eletrônico): 978-1-4842-3516-4. DOI: <https://doi.org/10.1007/978-1-4842-3516-4> . Número de controle da Biblioteca do Congresso: 2018938097.
- MORAN, T. The Command Language Grammars: a representation for the user interface of interactive computer systems. International Journal of Man-Machine Studies, Academic Press, p. 3-50, 1981
- NEVES, L. A. P.; NETO, H. V.; GONZAGA, A. (Ed.). Avanços em Visão Computacional. 1.ed. Curitiba, PR: Omnipax, 2012. 406 p. ISBN 978-85-64619-09-8.
- PATTANAYAK, Santanu. Pro Deep Learning com TensorFlow: uma abordagem matemática para inteligência artificial avançada em Python . Bangalore, Karnataka, Índia, 2017. ISBN-13 (pbk): 978-1-4842-3095-4. ISBN-13 (eletrônico): 978-1-4842-3096-1. DOI: <https://doi.org/10.1007/978-1-4842-3096-1> .
- RUSSELL, Stuart J.; NORVIG, Peter. Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited, 2016
- WANG, F. et al. Additive margin softmax for face verification. IEEE Signal Processing Letters, v. 25, n. 7, p. 926-930, July 2018

APÊNDICES

Apêndice A. Códigos, modelo, base de dados
GitHub - <https://github.com/MariaLimaS/IHC>

Apêndice B. Base de dados(GoogleDrive)
https://drive.google.com/drive/folders/1co_rfmpauhB2VnIchFiDypKljIsla_8s?usp=share_link

Apêndice C. Vídeos
Youtube - <https://youtu.be/-kDSHanIZq8?si=Nb4IW7gcuC1qkD7S>