

Otimização do Reconhecimento de Emoções com Ajuste de Hiperparâmetros e Balanceamento de Dados com Data Augmentation

Lara Toledo Cordeiro Ottoni
Departamento de Automação Industrial
Instituto Federal de Minas Gerais (IFMG)
Ouro Preto, Brasil
lara.ottoni@ifmg.edu.br

Jés de Jesus Fiais Cerqueira
Departamento de Engenharia Elétrica e Computação
Universidade Federal da Bahia (UFBA)
Salvador, Brasil
jes@ufba.edu.br

Abstract—Facial expressions play a crucial role in non-verbal communication, influencing social interactions and collaboration between humans and robots. This study focuses on optimizing facial expression emotion recognition by exploring different combinations of optimizers, learning rates, data augmentation techniques, and neural network architectures. Using the FER2013 dataset, we evaluated the use of data augmentation for data balancing and six different CNN architectures. Our results indicate that the best-performing configuration includes the Adam optimizer with a learning rate of 0.001, the use of data augmentation, and a Base CNN architecture, achieving an accuracy of 77.83%.

Index Terms—Emotion recognition, Facial expressions, CNN, Deep learning.

I. INTRODUÇÃO

As expressões faciais desempenham um papel crucial na comunicação, servindo como meio não verbal de demonstrar emoções. A capacidade de reconhecer essas emoções de forma precisa desempenha um papel importante na manutenção de conexões interpessoais e na colaboração em sociedade. O rosto humano é composto por inúmeras dinâmicas de estados emocionais, incluindo desde felicidade e tristeza até raiva, medo, surpresa e desgosto [1].

Essas expressões faciais são o resultado de complexas interações entre os músculos faciais, resultando em variações perceptíveis na forma e na aparência geral do rosto. O Reconhecimento de Emoções Faciais (*Facial Emotion Recognition* - FER), é uma área de estudo que reúne especialistas em visão computacional, computação afetiva, interação humano-computador e comportamento humano. Seu foco está na previsão de emoções por meio da análise de expressões faciais em imagens ou vídeos [2].

Nos últimos anos, avanços significativos têm sido feitos no desenvolvimento de técnicas para o FER. Com o surgimento das técnicas de aprendizado profundo, especialmente as redes neurais convolucionais, o campo experimentou uma transformação significativa. Os modelos de aprendizado profundo demonstraram uma capacidade excepcional de aprender automaticamente características diretamente das imagens fa-

ciais, resultando em uma notável melhoria na precisão das tarefas de reconhecimento de emoções [3].

Nos modelos de aprendizado profundo, a importância de ajustar os hiperparâmetros, como otimizador e taxa de aprendizado, não pode ser subestimada. Esses hiperparâmetros desempenham um papel relevante na eficiência do treinamento do modelo, afetando diretamente sua capacidade de convergir para uma solução ótima. Além disso, o uso de técnicas de *data augmentation* (DA) é fundamental para enriquecer o conjunto de dados disponível para o modelo. Ao gerar variações sintéticas dos dados, o DA não apenas aumenta a quantidade de dados disponíveis para treinamento, mas também auxilia na criação de um modelo mais robusto e em uma maior capacidade de generalização. Por fim, a investigação da arquitetura neural é crucial para encontrar a estrutura mais adequada para o problema em questão.

Nesse sentido, este artigo tem como objetivo elaborar uma metodologia que busca otimizar o reconhecimento de emoções da expressão facial. Neste sentido, são investigadas as melhores combinações entre otimizador, taxa de aprendizado, *data augmentation* e arquitetura neural.

Este trabalho está estruturado da seguinte forma: A seção II apresenta uma revisão dos trabalhos na área FER que realizam investigações para encontrar as melhores configurações. A seção III define a abordagem completa para obtenção dos resultados para base de dados FER2013. Os resultados são apresentados e discutidos na seção IV. E por fim, na seção V é realizado a conclusão do trabalho.

II. TRABALHOS RELACIONADOS

O domínio da pesquisa em reconhecimento de emoções faciais testemunhou numerosas inovações e avanços significativos ao longo do tempo. Em particular, as técnicas avançadas de aprendizado profundo têm impulsionado progressos substanciais nesse campo. A aplicação desses algoritmos tem suscitado a demanda por investigar e otimizar configurações visando o melhor desempenho do modelo.

Esta seção fornece uma análise dos estudos no âmbito do reconhecimento de emoções faciais. Neste contexto, foi

examinado uma variedade de trabalhos recentes que se concentram em explorar a melhor combinação das configurações do FER. Esses artigos realizam comparações de pelo menos um parâmetro, incluindo o otimizador, a taxa de aprendizado, *data augmentation* e a arquitetura neural. Na Tabela I, são apresentados os artigos e as configurações investigadas em seus estudos.

TABLE I
TRABALHO RELACIONADO NA ÁREA FER.

Artigo	Otimiz.	Taxa Aprend.	DA	Arq. Neural
[4]	✓	✓	-	✓
[5]	-	-	-	✓
[6]	-	-	-	✓
[7]	-	-	-	✓
[8]	-	-	-	✓
[9]	-	-	-	✓
[10]	-	-	-	✓
[3]	-	-	-	✓
[11]	-	-	-	✓
[12]	-	-	✓	✓
Proposta	✓	✓	✓	✓

No artigo de [4], uma comparação entre otimizadores é realizada, abrangendo tanto o Adam quanto o SGD. Além disso, foram analisadas as diferentes taxas de aprendizado, variando de 0,01, 0,001 a 0,005. Adicionalmente, houve uma avaliação comparativa das arquiteturas neurais da CNN. Os autores empregaram técnicas de *data augmentation*, como espelhamento horizontal, zoom, rotação e deslocamento em largura e altura. Contudo, é importante destacar que o efeito do DA não foi objeto de comparação.

No estudo conduzido por [5] e [6], são comparadas exclusivamente arquiteturas de redes neurais convolucionais. Em [5], o otimizador empregado foi o SGDM, utilizando uma taxa de aprendizado variável, juntamente com técnicas DA, como deslocamento em largura e altura, além de espelhamento horizontal e rotação. Enquanto isso, no estudo de [6], foi adotado o SGD com taxa de aprendizado variável, combinado com o DA de espelhamento e rotação.

Já no artigo de [7], o otimizador Adam foi empregado com uma taxa de aprendizado variável. Quanto ao DA, embora os autores mencionem várias técnicas, elas não são especificadas. O estudo se concentra exclusivamente na comparação de diferentes arquiteturas neurais.

Nos estudos conduzidos por [8] e [9], ambos abordam a comparação de arquiteturas neurais. Entretanto, nenhum dos dois artigos faz uso da técnica de *data augmentation*. No trabalho de [8], os autores optaram pelo otimizador Adam com uma taxa de aprendizado fixa de 0,001, enquanto no estudo de [9], não há descrição do otimizador e da taxa de aprendizado utilizados.

No artigo de [10], foi empregado o otimizador Adam com uma taxa de aprendizado variável. Para o *data augmentation*, foram aplicadas técnicas de espelhamento, rotação e corte aleatório. O estudo também compara arquiteturas que empregam o algoritmo CNN com mecanismo de atenção. Já o artigo de [3], não apresenta informações sobre o otimizador, taxa

de aprendizado e DA, mas também realiza uma comparação entre arquiteturas neurais utilizando CNN com mecanismo de atenção.

No estudo conduzido por [11], foi empregado o otimizador SGD com uma taxa de aprendizado variável. Quanto ao DA, foram empregadas transformações que incluem translação, rotação, espelhamento e zoom. O estudo realizou comparações apenas com relação à arquitetura neural da CNN.

E por fim, artigo realizado por [12] adota o otimizador Adam com uma taxa de aprendizado constante de 0,001. Nesse trabalho, os autores exploram técnicas aprimoradas de DA, incluindo espelhamento horizontal e rotação; a investigação abrange a variação das taxas dessas técnicas em quatro modelos distintos. Além do *data augmentation*, os autores também comparam quatro arquiteturas neurais diferentes.

A análise realizada evidencia uma forte tendência entre os estudos amostrados, dos quais 80% se concentram na investigação das arquiteturas neurais. Além disso, é possível observar que os trabalhos apontados pela Tabela I demonstram uma lacuna no que se refere a análise comparativa entre otimizador, taxa de aprendizado e *data augmentation*.

Diante disso, este trabalho tem como motivação atender a lacuna apontada pelos trabalhos citados anteriormente. Desta forma, a principal contribuição deste artigo é a elaboração de uma metodologia sistemática que investiga a combinação entre técnicas de *data augmentation*, otimizador, taxa de aprendizado e arquitetura neural para o reconhecimento de emoção facial.

III. METODOLOGIA PROPOSTA

Diante das considerações apresentadas na Seção II e das lacunas identificadas, foi desenvolvida uma metodologia para aprimorar as combinações do módulo de reconhecimento de emoções da expressão facial. Nesse contexto, serão avaliados técnicas de *data augmentation*, otimizadores, taxas de aprendizado e arquiteturas neurais.

A sequência das etapas da metodologia proposta é ilustrada na Figura 1. Para conduzir esta análise, será utilizada a base de dados FER2013, amplamente empregada na literatura para o reconhecimento de emoções na expressão facial.

A primeira etapa da metodologia será realizada uma análise sobre a eficácia das técnicas de aumento de dados (*data augmentation*) para o balanceamento do dataset. Na segunda etapa, consiste na investigação do otimizador e da taxa de aprendizado. Por fim, serão avaliados algoritmos de aprendizado profundo, Rede Neural Convolucional (CNN). A seguir, serão fornecidas informações mais detalhadas sobre cada uma dessas etapas.

A. Bases de Dados

A base de dados FER2013 é um conjunto de dados de código aberto carregado publicamente para uma competição *Kaggle*. Este conjunto de dados consiste em 35887 fotos de rosto em escala de cinza, 48×48 pixels com as sete emoções diferentes. O conjunto de dados foi definido como acessível a todos após a conclusão da competição. Na Figura 2, pode-se

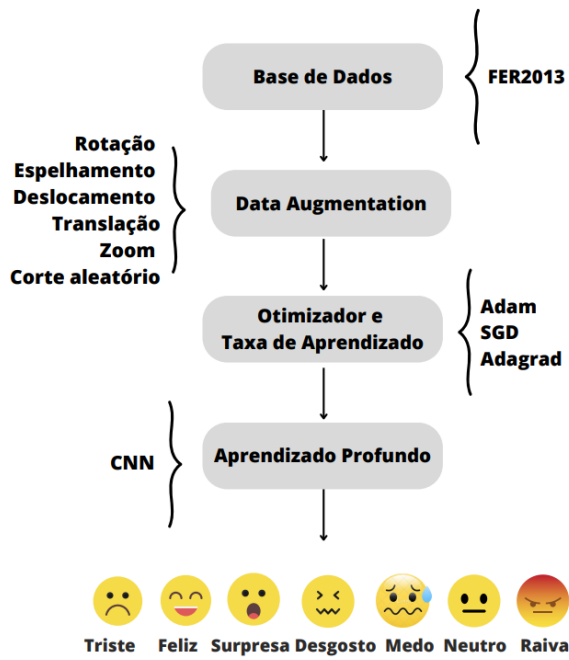


Fig. 1. Metodologia proposta para otimizar o reconhecimento de emoções da expressão facial

visualizar algumas das imagens contidas no conjunto FER2013 para as sete emoções: feliz, surpresa, desgosto, raiva, triste, neutro e medo. Com o total de 35887, o conjunto de dados FER2013 é desbalanceado, possuindo 4953 imagens de raiva, 547 de desgosto, 5121 de medo, 8989 de feliz, 6077 de neutro, 4002 de triste e 6198 de surpresa.



Fig. 2. Exemplo das imagens contidas na base de dados FER2013.

B. Passo 1: Otimização da data augmentation

A aplicação de técnicas de *data augmentation* é crucial no campo da visão computacional, especialmente em tarefas de reconhecimento de imagem, onde a disponibilidade de dados e a diversidade é essencial para treinar modelos eficazes [4]. Na Figura 3, é possível observar algumas das técnicas de *data augmentation* utilizadas.



Fig. 3. Exemplo de operações de *data augmentation*. Fonte da imagem Original: banco de dados WSEFEP [13].

Uma das técnicas mais simples e frequentemente empregadas é a rotação de imagens, Figura 3(a), na qual as amostras são giradas em ângulos específicos, permitindo ao modelo aprender a reconhecer objetos em diferentes orientações. Essa variação é essencial para garantir que o modelo seja robusto e capaz de lidar com objetos apresentados de maneiras diferentes nas imagens [14].

Outras técnicas importantes são o deslocamento em largura (Figura 3(b)), deslocamento em altura (Figura 3(c)), o espelhamento horizontal (Figura 3(d)) e o espelhamento vertical (Figura 3(e)) que reflete a imagem em relação a um eixo, introduzindo variações na posição dos objetos. Isso é fundamental para ensinar o modelo a reconhecer objetos independentemente da sua orientação espacial, o que é especialmente relevante em tarefas como detecção de objetos em que a posição do objeto pode variar consideravelmente [14].

Além disso, técnicas como cisalhamento (Figura 3(g)) e zoom (Figura 3(h)) também são amplamente utilizadas para introduzir variações na posição e escala dos objetos nas imagens, ajudando a melhorar a capacidade do modelo de generalizar para diferentes cenários. Técnicas como o ajuste de brilho e contraste, assim como a adição de ruído, são

formas de *data augmentation* utilizadas com menor frequência em comparação às demais, embora contribuam para tornar o modelo mais robusto a variações de iluminação e interferências externas. [14].

A primeira etapa da metodologia consiste na análise do uso de *data augmentation* para o balanceamento da base de dados FER2013. Para isso, foram testadas utilizando diferentes operações de aumento de dados, utilizando a função “ImageDataGenerator” do *framework* Keras. A seguir, estão descritos as operações utilizadas neste trabalho e seus respectivos parâmetros.

Rotação: 40%
Deslocamento em Largura: 0,2
Deslocamento em Altura: 0,2
Espelhamento Vertical: *true*
Espelhamento Horizontal: *true*
Cisalhamento: 0,8
Zoom: 20%

Com o objetivo de balancear a base de dados FER2013, investigou-se se as operações de *data augmentation*, aplicadas em combinação, seriam mais eficazes do que utilizar a base em sua forma desbalanceada. A acurácia sobre o conjunto de validação foi adotada como critério principal para avaliar o desempenho do modelo.

C. Passo 2: Ajuste dos otimizadores e taxa de aprendizado

Dada a importância desses parâmetros para o modelo CNN, a segunda investigação do método sistemático proposto visa determinar o melhor otimizador e valor da taxa de aprendizado para cada conjunto de dados utilizado. Para isso, serão considerados os principais otimizadores da literatura, como Adam, SGD e Adagrad, juntamente com valores de aprendizado comuns em estudos de reconhecimento de emoções através de expressões faciais (0,01 e 0,001).

D. Passo 3: Investigação da Arquitetura Neural

O terceiro passo da metodologia proposta consiste na investigação da arquitetura neural que seja capaz de gerar um bom desempenho na tarefa de reconhecimento de emoção facial. Para isso, foram explorados algoritmos baseados em Redes Neurais Convolucionais (CNNs). A arquitetura base (Modelo 1) adotada nos experimentos das etapas 1 e 2 é apresentada na Tabela II, sendo baseada na proposta descrita em [14].

Para comparar o desempenho da arquitetura base (Tabela II), foram utilizadas mais quatro arquiteturas CNNs. A arquitetura de Modelo 2, Tabela III, foi elaborada com o objetivo de investigar como uma rede com maior profundidade se comportaria com a base de dados utilizada [14].

O Modelo 3 (Tabela IV) tem o objetivo de avaliar o desempenho da substituição da camada de flatten por uma camada de Global Average Pooling. Já o Modelo 4, Tabela V, inverte a lógica comum de redes profundas, iniciando com um número alto de filtros e reduzindo progressivamente nas camadas subsequentes. O objetivo é verificar como a rede

TABLE II
ARQUITETURA DA CNN MODELO 1 - BASE

Camada	Parâmetros/Filtros	Ativação
Conv2D	32 filtros, 3×3	ReLU
BatchNormalization	—	—
MaxPooling2D	2×2	—
Conv2D	64 filtros, 3×3	ReLU
BatchNormalization	—	—
MaxPooling2D	2×2	—
Conv2D	128 filtros, 3×3	ReLU
BatchNormalization	—	—
MaxPooling2D	2×2	—
Flatten	—	—
Dropout	0,5	—
Dense	256 unidades	ReLU
Dropout	0,5	—
Dense (Saída)	7 unidades	Softmax

TABLE III
ARQUITETURA DA CNN - MODELO 2

Camada	Parâmetros/Filtros	Ativação
Conv2D	32 filtros, 3×3	ReLU
BatchNormalization	—	—
Conv2D	32 filtros, 3×3	ReLU
BatchNormalization	—	—
MaxPooling2D	2×2	—
Dropout	0,25	—
Conv2D	64 filtros, 3×3	ReLU
BatchNormalization	—	—
Conv2D	64 filtros, 3×3	ReLU
BatchNormalization	—	—
MaxPooling2D	2×2	—
Dropout	0,25	—
Conv2D	128 filtros, 3×3	ReLU
BatchNormalization	—	—
Conv2D	128 filtros, 3×3	ReLU
BatchNormalization	—	—
MaxPooling2D	2×2	—
Dropout	0,25	—
Flatten	—	—
Dense	512 unidades	ReLU
Dropout	0,5	—
Dense (Saída)	7 unidades	Softmax

TABLE IV
ARQUITETURA DA CNN - MODELO 3

Camada	Parâmetros/Filtros	Ativação
Conv2D	64 filtros, 3×3	ReLU
BatchNormalization	—	—
MaxPooling2D	2×2	—
Conv2D	128 filtros, 3×3	ReLU
BatchNormalization	—	—
MaxPooling2D	2×2	—
Conv2D	256 filtros, 3×3	ReLU
BatchNormalization	—	—
GlobalAveragePooling2D	—	—
Dense (Saída)	7 unidades	Softmax

se comporta com uma extração inicial de grande volume de características, seguida de uma compressão gradual [14].

E por fim, o Modelo 5 (Tabela VI) foi inspirada na arquitetura VGG, essa configuração testa o efeito de blocos convolucionais duplos com filtros constantes dentro de cada bloco, seguidos por pooling. A proposta busca analisar o desempenho de uma arquitetura conhecida, adaptada para imagens menores

TABLE V
ARQUITETURA DA CNN - MODELO 4

Camada	Parâmetros/Filtros	Ativação
Conv2D	128 filtros, 3×3	ReLU
BatchNormalization	—	—
MaxPooling2D	2×2	—
Conv2D	64 filtros, 3×3	ReLU
BatchNormalization	—	—
MaxPooling2D	2×2	—
Conv2D	32 filtros, 3×3	ReLU
BatchNormalization	—	—
Flatten	—	—
Dense	128 unidades	ReLU
Dropout	0,4	—
Dense (Saída)	7 unidades	Softmax

TABLE VI
ARQUITETURA DA CNN - MODELO 5

Camada	Parâmetros/Filtros	Ativação
Conv2D	64 filtros, 3×3	ReLU
Conv2D	64 filtros, 3×3	ReLU
MaxPooling2D	2×2	—
Dropout	0,25	—
Conv2D	128 filtros, 3×3	ReLU
Conv2D	128 filtros, 3×3	ReLU
MaxPooling2D	2×2	—
Dropout	0,25	—
Flatten	—	—
Dense	512 unidades	ReLU
Dropout	0,5	—
Dense (Saída)	7 unidades	Softmax

(48x48), com menos camadas e menos parâmetros do que a VGG original.

Para a investigação proposta, os dados foram divididos em conjuntos de treinamento (80%), validação (10%) e teste (10%). O treinamento foi realizado com *batch size* igual a 64, ao longo de 100 épocas. A função perda utilizada foi a Categorical Crossentropy.

IV. RESULTADOS

Nesta seção serão apresentados os resultados obtidos seguindo a metodologia descrita na Seção III. Para realizar os experimentos utilizou-se um notebook com sistema operacional Windows 11, processador Intel i5 1135G7 2.40GHz, memória RAM de 8 GB e uma GPU Nvidia GeForce MX350 com 2GB VRAM. O ambiente de desenvolvimento utilizado neste trabalho é composto pela IDE Jupyter, associada a linguagem Python.

Conforme descrito na Seção III a primeira etapa da investigação metodológica consiste em utilizar o *data augmentation*. Na Tabela VII é possível observar a influência de se aplicar a técnica ao *deep learning* para o balanceamento da base FER2013.

TABLE VII
RESULTADO - PASSO 1.

Data Augmentation	Acurácia
Base desbalanceada	60,77%
Base balanceada	77,83%

Como pode ser observado, o conjunto de dados balanceado apresentou um desempenho superior em comparação ao conjunto desbalanceado. Após o balanceamento, todas as sete classes passaram a conter 8.989 imagens cada, totalizando 62.923 amostras. Esse resultado evidencia que o balanceamento das classes contribui significativamente para a melhoria da capacidade de generalização do modelo, evitando o viés em favor das classes com mais imagens.

A segunda etapa da metodologia, consiste em encontrar o melhor conjunto de otimizador e taxa de aprendizado para o modelo da rede neural convolucional que classifica as emoções da FER2013. Neste sentido, foram testados os otimizadores Adam, SGD e Adagrad juntamente com as taxas de aprendizado 0,01 e 0,001. Os resultados estão descritos na Tabela VIII.

TABLE VIII
RESULTADO - PASSO 2.

Taxa de Aprendizado	Adam	SGD	Adagrad
0,01	56,90%	73,62%	75,66%
0,001	77,83%	68,17%	66,45%

Como pode ser visto na Tabela VIII, o otimizador Adam em conjunto com a taxa de aprendizado 0,001 obtiveram o melhor valor de acurácia. Desta forma, nas próximas etapas da metodologia serão utilizados o conjunto Adam 0,001.

E por fim, a última etapa da metodologia proposta é a investigação da arquitetura neural. Nesse trabalho utilizou-se seis arquiteturas, conforme mostrado na Seção III, uma arquitetura CNN base e mais cinco arquiteturas CNNs propostas para investigação. Na Tabela IX é possível verificar os resultados de acurácia obtidos em cada arquitetura neural.

TABLE IX
RESULTADO - PASSO 3.

Arquitetura	Acurácia (%)
CNN Modelo 1 - Base	77,83
CNN Modelo 2	76,55
CNN Modelo 3	73,24
CNN Modelo 4	70,63
CNN Modelo 5	76,62

Desta forma, segundo a metodologia proposta, a melhor combinação para o reconhecimento de emoções pela expressão facial utilizando a base de dados FER2013 é com o otimizador Adam, taxa de aprendizado 0,001, usando a técnica de *data augmentation* para balancear a base de dados FER2013 e com a arquitetura neural CNN Modelo 1 - Base. Essa combinação gerou um histórico de acurácia e perda mostrado na Figura 4 e valores de precisão, recall e f-score mostrados na Tabela X.

A. Comparação dos Resultados

Esta seção se destina a realizar uma análise comparativa dos resultados obtidos em relação aos estudos apresentados na Tabela I. Com esse propósito, a Tabela XI é disponibilizada, oferecendo uma visão das métricas de acurácia encontradas em cada um dos artigos selecionados.

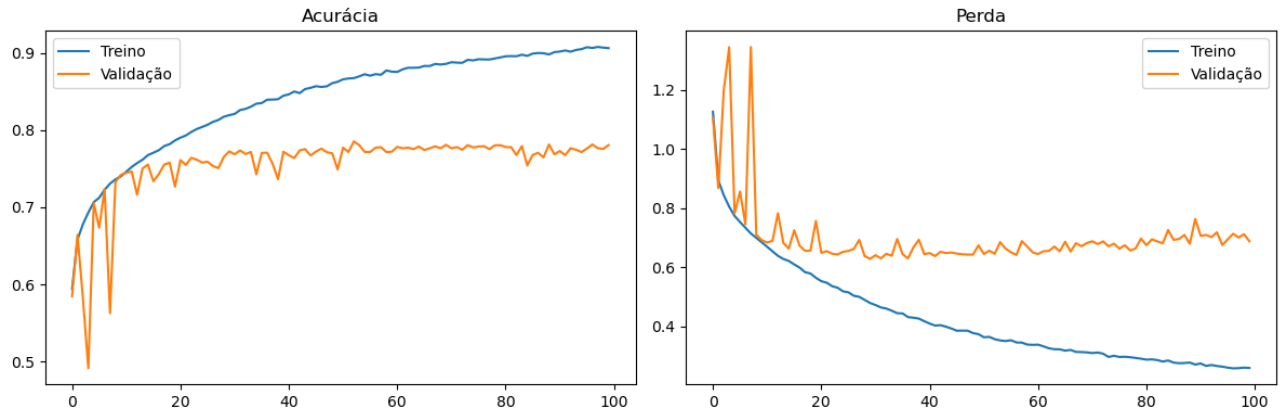


Fig. 4. Gráficos do histórico de treinamento da acurácia e perda

TABLE X
VALORES EM PORCENTAGEM DA PRECISÃO, RECALL E F-SCORE PARA A
BASE FER2013.

Emoção	Precisão (%)	Recall (%)	F-Score (%)
Desgosto	81,00	65,0	72,0
Feliz	81,0	82,0	82,0
Medo	75,0	63,0	69,0
Neutro	68,0	88,0	77,0
Raiva	88,0	85,0	87,0
Surpresa	62,0	74,0	68,0
Triste	75,0	81,0	77,0
Acurácia (%)	-	-	78,0

TABLE XI
COMPARAÇÃO ENTRE OS VALORES DE ACURÁCIA PARA O
RECONHECIMENTO DE EMOÇÃO PELA EXPRESSÃO FACIAL QUE UTILIZAM
A BASE DE DADOS FER2013.

Trabalhos	Acurácia (%)
Proposto	77,83
[4]	75,15
[5]	66,60
[6]	61,05
[7]	83,00
[8]	73,31
[9]	72,35
[10]	74,00
[3]	71,38
[11]	74,09
[12]	69,57

A partir dos dados da Tabela XI, observa-se que as acurácias reportadas nos estudos variam entre aproximadamente 61% e 83%. Essa variação é coerente com a dificuldade amplamente reconhecida na literatura em relação à base FER2013, conhecida por sua complexidade. Nesse contexto, a metodologia proposta, que alcançou uma acurácia de 77,83%, posiciona-se de forma competitiva dentro da faixa observada, evidenciando a consistência e a relevância dos resultados obtidos.

V. CONCLUSÃO

Este trabalho teve como objetivo propor uma metodologia sistemática para o ajuste dos parâmetros de um modelo de reconhecimento de emoções faciais. Para isso, foi desenvolvida

uma sequência de etapas voltadas o uso do *data augmentation* para balancear os dados, escolha do otimizador, definição da taxa de aprendizado e da arquitetura da rede neural convolucional.

Utilizando a base de dados FER2013, conhecida por sua complexidade e desbalanceamento, foi alcançada uma acurácia de 77,83%. Esse valor está em conformidade com os resultados reportados na literatura. Além disso, o balanceamento das classes mostrou-se um fator essencial para o aprimoramento do desempenho, contribuindo para a redução de viés e o aumento da capacidade de generalização do modelo.

Como trabalhos futuros, pretende-se realizar novos testes com diferentes bases de dados, explorar o uso de técnicas mais avançadas de *data augmentation*. Além disso, pretende-se comparar o desempenho com arquiteturas mais recentes, como redes pré-treinadas.

AGRADECIMENTOS

Os autores agradecem a UFBA, CAPES (Código de Financiamento 001), FAPESB (084.0508.2022.0000 201-09) e IFMG/Ouro Preto.

REFERENCES

- [1] C. Dewi, L. S. Gunawan, S. G. Hastoko, and H. J. Christanto, "Real-time facial expression recognition: Advances, challenges, and future directions," *Vietnam Journal of Computer Science*, pp. 1–27, 2023.
- [2] L. T. C. Ottoni and J. J. F. Cerqueira, "A systematic review of human-robot interaction: the use of emotions and the evaluation of their performance," *International Journal of Social Robotics*, pp. 1–20, 2024.
- [3] Y. Wu, "Facial expression recognition in classroom environment based on attention mechanism," in *Advances in Artificial Intelligence, Big Data and Algorithms*. IOS Press, 2023, pp. 689–695.
- [4] R. Lawpanom, W. Songpan, and J. Kaewyotha, "Advancing facial expression recognition in online learning education using a homogeneous ensemble convolutional neural network approach," *Applied Sciences*, vol. 14, no. 3, p. 1156, 2024.
- [5] G. K. Sahoo, S. K. Das, and P. Singh, "Performance comparison of facial emotion recognition: a transfer learning-based driver assistance framework for in-vehicle applications," *Circuits, Systems, and Signal Processing*, vol. 42, no. 7, pp. 4292–4319, 2023.
- [6] T.-D. Pham, M.-T. Duong, Q.-T. Ho, S. Lee, and M.-C. Hong, "Cnn-based facial expression recognition with simultaneous consideration of inter-class and intra-class variations," *Sensors*, vol. 23, no. 24, p. 9658, 2023.

- [7] R. Helaly, S. Messaoud, S. Bouaafia, M. A. Hajjaji, and A. Mtibaa, "Dtl-i-resnet18: Facial emotion recognition based on deep transfer learning and improved resnet18," *Signal, Image and Video Processing*, vol. 17, no. 6, pp. 2731–2744, 2023.
- [8] J. Kim and D. Lee, "Facial expression recognition robust to occlusion and to intra-similarity problem using relevant subsampling," *Sensors*, vol. 23, no. 5, p. 2619, 2023.
- [9] H. Li and H. Xu, "Deep reinforcement learning for robust emotional classification in facial expression recognition," *Knowledge-Based Systems*, vol. 204, p. 106172, 2020.
- [10] Z. Song, "Facial expression emotion recognition model integrating philosophy and machine learning theory," *Frontiers in Psychology*, vol. 12, p. 759485, 2021.
- [11] H.-D. Nguyen, S. Yeom, G.-S. Lee, H.-J. Yang, I.-S. Na, and S.-H. Kim, "Facial emotion recognition using an ensemble of multi-level convolutional neural networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 11, p. 1940015, 2019.
- [12] J. D. Bodapati, U. Srilakshmi, and N. Veeranjanyulu, "Fernet: a deep cnn architecture for facial expression recognition in the wild," *Journal of The institution of engineers (India): series B*, vol. 103, no. 2, pp. 439–448, 2022.
- [13] M. Olszanowski, G. Pochwatko, K. Kuklinski, M. Scibor-Rylski, P. Lewinski, and R. K. Ohme, "Warsaw set of emotional facial expression pictures: a validation study of facial display photographs," *Frontiers in psychology*, vol. 5, p. 1516, 2015.
- [14] F. Chollet, *Deep learning with Python*. Simon and Schuster, 2021.