

# Reconhecimento de Emoções com Redes Neurais para Apoio à Segurança da Mulher

Mariana de Oliveira Lopes  
Departamento de Computação  
Universidade Federal de Ouro Preto  
Ouro Preto, Brasil  
mariana.ol@aluno.ufop.edu.br

**Resumo**—A segurança das mulheres em espaços públicos constitui um desafio global urgente, evidenciado pelo aumento de casos de violência de gênero. Neste trabalho, propõe-se desenvolver um sistema de reconhecimento de padrões de emoções faciais, visando classificar e identificar automaticamente indícios de medo, tristeza e raiva em situações potencialmente adversas. Serão empregadas redes neurais convolucionais profundas baseadas na arquitetura ResNet50, aplicando-se técnicas de *fine-tuning* e *data augmentation* em imagens de rostos extraídas dos conjuntos FER2013, RAF-DB e KDEF. O modelo deverá ser treinado para classificar três categorias emocionais: medo e tristeza em rostos femininos e raiva em rostos masculinos. A avaliação de desempenho contemplará métricas de acurácia, precisão, *recall* e *F1-score*, com ênfase na robustez e na capacidade de generalização do sistema. Futuramente, espera-se que esta abordagem contribua para o desenvolvimento de sistemas de monitoramento inteligentes voltados à proteção de mulheres em contextos de vulnerabilidade.

**Palavras-chave**—Emoções Faciais, Reconhecimento de Padrões, Redes Neurais Convolucionais, Segurança da Mulher.

## I. INTRODUÇÃO

A segurança das mulheres é uma preocupação global urgente, refletida em dados alarmantes que evidenciam a persistência da violência de gênero. Em 2023, o Brasil registrou 1.463 feminicídios, representando uma taxa de 1,4 mortes para cada 100 mil mulheres, o que indica um aumento de 1,6% em relação ao ano anterior [1]. Além disso, o assédio e a importunação sexual continuam sendo problemas recorrentes e opressores. Essa realidade afeta diretamente a qualidade de vida e a participação plena das mulheres na sociedade, sendo considerada uma barreira estrutural ao desenvolvimento igualitário [2, 3].

O avanço das tecnologias digitais, especialmente no campo da Inteligência Artificial (IA), tem impulsionado o desenvolvimento de soluções inovadoras para enfrentar problemas sociais complexos, como a violência de gênero. A IA permite a construção de sistemas capazes de identificar e interpretar padrões complexos em grandes volumes de dados, sendo amplamente utilizada em aplicações de monitoramento e detecção de riscos [4]. Dentre suas abordagens, o Aprendizado de

Máquina (AM) destaca-se por sua capacidade de reconhecer padrões sutis, como microexpressões faciais, por meio do processamento de dados visuais [5]. Neste trabalho, técnicas de AM são aplicadas a imagens de rostos humanos com o objetivo de identificar emoções como medo, tristeza e raiva — sentimentos frequentemente associados a situações de assédio ou ameaça [6]. O reconhecimento automatizado desses padrões emocionais pode ser um importante recurso em sistemas de alerta precoce, contribuindo de forma preventiva para a proteção de mulheres em contextos de vulnerabilidade.

Para alcançar uma identificação precisa de emoções faciais em contextos de risco, este trabalho adota uma abordagem baseada em redes neurais convolucionais profundas, utilizando a arquitetura ResNet50 combinada com técnicas de *fine-tuning* e *data augmentation*. A ResNet50 é amplamente reconhecida por sua capacidade de extrair características discriminativas de imagens complexas, devido à sua profundidade e ao uso de conexões residuais que mitigam problemas como o desaparecimento do gradiente [7]. Essa arquitetura tem demonstrado desempenho superior em tarefas de reconhecimento de expressões faciais, especialmente quando adaptada por meio de *fine-tuning* em conjuntos de dados específicos [8]. Além disso, a aplicação de técnicas de *data augmentation*, como ajustes de saturação e brilho, contribui para a ampliação e diversificação do conjunto de dados de treinamento, reduzindo o risco de overfitting e melhorando a generalização do modelo [9]. A combinação dessas estratégias permite a construção de um sistema robusto e eficiente para a detecção automatizada de emoções associadas a situações de vulnerabilidade feminina.

Neste contexto, o presente trabalho tem como objetivo construir um sistema de classificação de emoções faciais a partir de imagens, com foco na identificação de sinais que possam indicar potenciais situações de assédio. Para isso, serão consideradas duas categorias emocionais: uma que reúne expressões de medo e tristeza em rostos femininos, e outra correspondente à expressão de raiva em rostos masculinos. O problema é formulado como uma tarefa de classificação binária, na qual o modelo deverá, a partir de uma imagem de rosto, distinguir entre essas duas classes pré-estabelecidas.

## II. OBJETIVOS

### A. Objetivo Geral

Desenvolver um modelo de reconhecimento de emoções baseado em redes neurais convolucionais, com foco na identificação das expressões faciais de **medo, tristeza e raiva**, visando à detecção precoce de possíveis situações de assédio ou ameaça contra mulheres em espaços públicos.

### B. Objetivos Específicos

- Investigar e selecionar subconjuntos relevantes das bases de dados **FER2013**, **RAF-DB** e **KDEF**, contendo imagens faciais rotuladas com as emoções de medo, tristeza e raiva, assegurando diversidade de gênero e qualidade dos dados.
- Aplicar técnicas de *data augmentation* para ampliar e balancear os dados de treinamento, aumentando a robustez e a generalização do modelo.
- Implementar um modelo de rede neural utilizando a arquitetura **ResNet50**, realizando *fine-tuning* para adaptar a rede ao domínio específico de emoções humanas.
- Avaliar o desempenho do modelo por meio de métricas como acurácia, precisão, *recall* e F1-score, com especial atenção às classes que indicam risco.
- Discutir o potencial uso do modelo em sistemas de monitoramento inteligentes voltados à segurança pública e à proteção de mulheres em situações de vulnerabilidade.

## III. TRABALHOS RELACIONADOS

O reconhecimento de emoções em imagens e em sinais de fala tem avançado significativamente com o uso de arquiteturas profundas de redes neurais. Nesta seção, revisamos cinco trabalhos fundamentais que embasam a nossa proposta.

### A. CNN-LSTM with Chebyshev Moments and SVM Classification

Jayaraman e Mahendran [10] propõem um framework híbrido para detecção de emoções faciais em motoristas, com o objetivo de aumentar a segurança viária ao identificar estados emocionais como estresse, cansaço e distração. A motivação decorre da necessidade de monitorar mudanças sutis na expressão facial que possam indicar riscos à condução.

Para isso, os autores utilizam a base FER2013, originalmente composta por sete classes emocionais. Após um rigoroso processo de seleção manual — baseado em qualidade de imagem, contraste e visibilidade das feições — o conjunto foi reduzido a 7 074 amostras, distribuídas em cinco categorias (*feliz, triste, zangado, medo e neutro*), garantindo um dataset mais balanceado e adequado à tarefa.

A arquitetura integra uma CNN baseada em *ResNet152*, que extrai características espaciais profundas por meio de conexões residuais (*skip connections*), atenuando o problema de gradiente desaparecendo, e uma camada LSTM (*Long Short-Term Memory*) para capturar a dinâmica temporal entre frames, essencial na distinção de transições expressivas sutis. Na etapa de pré-processamento, aplicam-se momentos de Chebyshev para reforçar a extração de atributos estatísticos,

e a robustez do modelo é assegurada por validação cruzada *K-Fold*. Finalmente, os vetores de características são encaminhados a um SVM (*Support Vector Machine*) com kernel RBF, implementado em múltiplas bibliotecas para validar a consistência dos resultados.

Durante o desenvolvimento, enfrentaram-se desafios como desbalanceamento de classes, variações de contraste e oclusões faciais. O balanceamento foi alcançado pela redução criteriosa do dataset, enquanto a normalização e filtragem de imagens mitigaram os efeitos de iluminação desigual e partes do rosto ocultas.

O modelo proposto atinge 82% de acurácia, 84% de *recall*, 92% de precisão, 87% de sensibilidade e *F1-Score* de 91%, superando significativamente as abordagens comparativas: CNN-SVM ( $\approx 63\%$  de acurácia) e ANN-LSTM ( $\approx 69\%$  de acurácia).

### B. An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks

Białek, Matiolański e Grega [11] apresentam uma solução voltada para aplicações práticas de reconhecimento de emoções — como teleconferências, entrevistas online e interfaces adaptativas — onde respostas emocionais precisas melhoram a interação humano-computador. Eles partem do desafio de detecção de expressões de baixa intensidade (“*micro-expressions*”), além do desbalanceamento de classes e de rótulos ruidosos na base FER2013.

A base FER2013, com suas sete emoções clássicas, serve de referência. Para lidar com classes raras, os autores criam cenários binários (uma rede por emoção, treinada para distinguir “emoção” vs. “outras”), cada um com 6000 amostras balanceadas por *data augmentation* e *undersampling*, assegurando distribuição uniforme durante o treino.

No aspecto arquitetural, propõem dois modelos *from scratch*: um de cinco camadas convolucionais e outro de seis camadas que substitui as camadas totalmente conectadas, reduzindo drasticamente o número de parâmetros e o risco de *overfitting*. Complementam com *Transfer Learning* usando VGG16 e ResNet50 pré-treinadas em ImageNet. Para cada configuração — os dois modelos customizados e as duas redes de transferência — treinam versões multiclases e binárias, totalizando 21 redes, com até 200 épocas nas CNNs customizadas e 100 nas de transferência.

Os experimentos, realizados em Google Colab com GPU Tesla T4 e implementados em Keras/TensorFlow, mostram que, isoladamente, a ResNet50 atinge 72,72% de acurácia no FER2013 original, enquanto os modelos customizados balanceados chegam a 69–71%, com latência de aproximadamente 0,62s para 1000 imagens. Ao combinar quatro redes de transferência (treinadas em conjuntos original e balanceado) em um ensemble, alcançam 75,06% de acurácia no FER2013 original e 76,90% no conjunto filtrado. Notavelmente, os classificadores binários chegam a mais de 90% para emoções como “feliz” e “nojo”, embora a concatenação de suas saídas para reconhecimento multiclases degrade o desempenho geral.

Este estudo demonstra que, mesmo com arquiteturas relativamente simples, a combinação de *data augmentation*, balanceamento binário e ensembles de modelos de transferência resulta em precisão competitiva, mantendo latência adequada para cenários em tempo real.

### C. Image-based Facial Emotion Recognition using Convolutional Neural Network on Emognition Dataset

Agung, Rifai e Wijayanto [12] partem da premissa de que as emoções humanas vão além das categorias básicas tradicionalmente reconhecidas por sistemas de FER (por exemplo, as seis ou sete emoções do FER2013 ou CK+), incluindo manifestações como *admiração* e *simpatia* que são cruciais em contextos de comunicação, marketing e interfaces adaptativas. A motivação central do trabalho é avaliar se redes convolucionais profundas conseguem diferenciar um conjunto ampliado de dez estados emocionais, extraídos de vídeos no dataset *Emognition*, convertidos para frames estáticos.

O dataset *Emognition* contém dez classes emocionais — *awe* (admiração), *enthusiasm* (entusiasmo), *gratitude* (gratidão), *contentment* (contentamento), *love* (amor), *optimism* (otimismo), *relief* (alívio), *sadness* (tristeza), *anger* (raiva) e *fear* (medo) — totalizando uma diversidade que ultrapassa as bases convencionais. Adicionalmente, o estudo avalia o desempenho das mesmas arquiteturas em bases clássicas (JAFPE e KDEF) para análise comparativa de generalização.

Metodologicamente, são comparadas duas abordagens principais: (i) *Transfer Learning*, empregando Inception-V3 (311 camadas) e MobileNet-V2 (154 camadas) pré-treinadas em ImageNet e refinadas (*fine-tuning*) nos frames do *Emognition*; (ii) *Full Learning*, em que uma CNN de arquitetura customizada é treinada do zero, com camada de convolução, pooling e camadas densas, seguindo boas práticas de normalização e regularização.

Os resultados destacam o poder do *fine-tuning*: o Inception-V3 ajustado atinge 96% de acurácia global e *F1-Score* médio de 0,95, superando claramente o MobileNet-V2 (89% de acurácia, *F1-Score* = 0,82) e a CNN *full learning* (87% de acurácia, *F1-Score* = 0,90). Além disso, o Inception-V3 demonstrou bom equilíbrio entre tempo de treinamento e desempenho de inferência, enquanto a abordagem *full learning*, embora mais simples e ligeiramente mais rápida na inferência, sofreu com a menor disponibilidade de dados por classe.

Em síntese, o trabalho confirma que arquiteturas profundas e pré-treinadas podem ser eficientemente adaptadas para conjuntos de emoções mais ricos, mantendo elevada precisão e robustez, o que abre caminho para sistemas de FER capazes de capturar nuances emocionais avançadas em aplicações reais.

### D. ResEmoteNet: Bridging Accuracy and Loss Reduction in Facial Emotion Recognition

Roy et al.[13] apresentam o ResEmoteNet, motivado pela crescente demanda por sistemas de reconhecimento de emoções faciais que funcionem de forma robusta em ambientes “in the wild” e que, simultaneamente, maximizem a acurácia e reduzam a perda de treinamento. Tais sistemas têm aplicação

em interfaces adaptativas, monitoramento de saúde mental e análise de comportamento em cenários de baixa qualidade de captura. O framework integra três blocos principais: um módulo CNN de três camadas convolucionais com Batch Normalization e Max-Pooling para extração hierárquica de padrões, blocos Squeeze-and-Excitation (SE) para atenção de canais que enfatiza características relevantes e suprime ruídos, e blocos residuais para mitigar o problema de gradiente desaparecendo em redes profundas. Ao final, aplica-se Adaptive Average Pooling para ajustar dinamicamente o tamanho do vetor de características antes da camada de classificação.

Para avaliação, os autores utilizaram três bases públicas de grande renome: FER2013, RAF-DB e AffectNet, cobrindo desde imagens em estúdio até cenários com variação de pose, iluminação e oclusões. O pré-processamento incluiu redimensionamento, normalização de pixels e aumento mínimo (flip horizontal), mantendo configuração uniforme de *batch size* = 16, 80 épocas de treino, função de custo cross-entropy e otimizador SGD com scheduler de learning rate.

Os resultados experimentais mostram que o ResEmoteNet supera o estado-da-arte em todos os três benchmarks, alcançando 79,79% de acurácia no FER2013, 94,76% no RAF-DB e 72,93% no AffectNet, com ganhos absolutos de 2,97%, 2,19% e 3,53% sobre os melhores métodos comparados. As matrizes de confusão evidenciam forte desempenho em classes fáceis como “happy” e “surprise”, ao mesmo tempo em que mantêm robustez em emoções menos nítidas. Esses resultados destacam a eficácia da combinação de atenção de canais e conexões residuais na melhoria simultânea de acurácia e estabilidade de treinamento.

### E. BRG-NeXt: Facial Affect Computing Using Adaptive Residual Networks With Bounded Gradient

Hasani et al. [14] propõem o BRG-NeXt, uma arquitetura de rede residual em que o atalho tradicional de identidade é substituído por uma função diferenciável de derivada limitada, com dois parâmetros treináveis ( $\alpha$  e  $\beta$ ). Essa modificação — o mapeamento adaptativo com gradiente limitado — melhora o fluxo de gradiente, evita problemas de vanishing/exploding e permite à rede capturar micro-expressões de forma mais eficiente, mantendo profundidade moderada e baixo custo computacional.

Para avaliar o modelo, os autores utilizam três bases de dados em ambiente natural: AffectNet, FER2013 e Affect-in-Wild, cobrindo tanto tarefas de classificação categórica quanto predição dimensional de valência e arousal.

No pré-processamento, as faces são recortadas e redimensionadas, seguidas de aumentos de dados (flip, variações de brilho e contraste), normalização, BatchNorm e ativação ELU. A arquitetura BRG-NeXt-50 (50 camadas, 3,1 M de parâmetros, 15 MFLOPs) é treinada com focal loss para o modelo categórico e MSE para o dimensional, usando ADAM (batch size 128, learning rate inicial de  $1e-4$ , decaimento periódico).

Em comparação com ResNet-50 e outros métodos *state-of-the-art*, o BRG-NeXt-50 atinge 68,50% de acurácia em

AffectNet e 71,53% em FER2013 no cenário categórico, além de RMSE de 0,2577 em AffectNet e 0,2882 em *Affect-in-Wild* no cenário dimensional, com elevados coeficientes de correlação e SAGR  $\approx 0,8$ . Esses resultados demonstram que o mapeamento adaptativo contribui significativamente para o reconhecimento de emoções em cenários não controlados.

#### F. Análise Crítica

Os cinco trabalhos revisados oferecem avanços importantes para o campo do reconhecimento de emoções, sendo particularmente úteis para fundamentar nossa proposta de um sistema voltado à segurança da mulher. A principal contribuição de Jayaraman e Mahendran [10] está na integração de CNN-LSTM com momentos de Chebyshev e SVM, demonstrando a eficácia de modelos híbridos em contextos com variabilidade emocional sutil — algo essencial para detectar sinais de medo ou tristeza em situações de risco.

Białek et al. [11] reforçam a importância do balanceamento de classes e do uso de classificadores binários, abordagem que pode ser útil ao restringirmos nosso escopo a emoções específicas, como medo, tristeza e raiva. A proposta de Agung et al. [12], ao utilizar um conjunto ampliado de emoções, evidencia a capacidade de redes profundas diferenciarem emoções mais complexas, o que pode ser explorado futuramente na personalização de respostas a diferentes contextos emocionais.

Roy et al. [13] demonstram que a combinação de atenção de canais com blocos residuais é eficaz para cenários não controlados, como ambientes públicos e com iluminação variável — características comuns aos contextos urbanos em que se espera aplicar nosso sistema. Por fim, Hasani et al. [14] introduzem um mecanismo adaptativo de residuais com gradiente limitado, mostrando-se promissor para capturar microexpressões, o que pode aumentar a sensibilidade do sistema em identificar emoções sutis de medo ou desconforto.

Essas contribuições apontam caminhos técnicos e metodológicos sólidos para a construção de um sistema robusto e sensível às nuances emocionais envolvidas em situações de potencial ameaça ou assédio. Para consolidar essa análise, a Tabela I apresenta um resumo comparativo dos principais aspectos dos trabalhos revisados, destacando as bases de dados utilizadas, as arquiteturas aplicadas e as principais técnicas e resultados alcançados. Essa comparação facilita a identificação de abordagens promissoras e limitações que podem orientar o desenvolvimento da solução proposta neste trabalho.

#### IV. BASES DE DADOS

Este trabalho utiliza três bases de dados amplamente reconhecidas para o treinamento do modelo de reconhecimento de emoções: FER2013 (*Facial Expression Recognition 2013 Dataset*), RAF-DB (*Real-world Affective Faces Database*) e KDEF Karolinska Directed Emotional Faces), ambas recorrentes na literatura e exploradas em estudos recentes [10, 11, 12, 13, 15, 17]. Cada uma delas possui características específicas que complementam o desenvolvimento e

TABLE I  
COMPARAÇÃO DOS PRINCIPAIS TRABALHOS

Artigo	Base de Dados	Arquitetura / Rede Neural	Técnicas e Resultados
CNN-LSTM with Chebyshev Moments and SVM Classification [10]	FER2013 (7 074 imagens)	ResNet152 + LSTM + SVM	Momentos de Chebyshev; K-fold; 82% acurácia, 92% precisão
An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks [11]	FER2013	CNN customizadas, VGG16, ResNet50, Ensemble	Filtragem e data augmentation; multiclass vs. binário; acurácia 70%
Image-based Facial Emotion Recognition using Convolutional Neural Network on Emognition Dataset [12]	Emognition (10 emoções)	Inception-V3, MobileNet-V2, CNN custom	Fine-tuning e Taguchi; data augmentation; 96% acurácia, F1-score 0,95
ResEmoteNet: Bridging Accuracy and Loss Reduction in Facial Emotion Recognition [13]	FER2013, RAF-DB, AffectNet	CNN 3 camadas convolucionais + Squeeze-and-Excitation Blocks + Residual Blocks + Adaptive Average Pooling	Data augmentation mínima; SGD com scheduler de learning rate; 79,79% acurácia (FER2013), 94,76% (RAF-DB), 72,93% (AffectNet)
BReG-NeXt: Facial Affect Computing Using Adaptive Residual Networks With Bounded Gradient [14]	AffectNet, FER2013, Affect-in-Wild	BReG-NeXt-50 (Adaptive Residual Network com atalho diferenciável)	Mapeamento adaptativo com derivada limitada; ADAM; focal loss/MSE; 68,50% acurácia (AffectNet), 71,53% (FER2013); RMSE 0,2577 (AffectNet), 0,2882 (Affect-in-Wild); SAGR $\approx 0,8$

a avaliação do modelo, abrangendo desde cenários controlados até situações do mundo real.

A base de dados *FER2013* foi criada para a competição do International Conference on Machine Learning (ICML) em 2013. Ela contém 35.887 imagens em escala de cinza, todas com resolução de 48x48 pixels (ver amostra na Figura 1), coletadas de maneira automática na internet. As imagens foram capturadas em diversas condições de iluminação, ângulos e expressões faciais. As emoções estão divididas em sete categorias principais: raiva, nojo, medo, felicidade, tristeza, surpresa e neutro [15]. Apesar de amplamente utilizada como *benchmark* na validação de modelos de aprendizado profundo, a FER2013 apresenta alguns desafios, como a presença de imagens com rotulagens imprecisas, o que pode influenciar os

resultados quando utilizada de forma isolada.



Fig. 1. FER2013. Fonte: [https://medium.com/@birdortyedi\\_23820/deep-learning-lab-episode-3-fer2013-c38f2e052280](https://medium.com/@birdortyedi_23820/deep-learning-lab-episode-3-fer2013-c38f2e052280).

A base de dados *RAF-DB* consiste em aproximadamente 30.000 imagens de rostos capturadas em cenários do mundo real (ver amostra na Figura 2). Esse conjunto de dados foi projetado para refletir os desafios presentes em situações reais, como variações de pose, iluminação, oclusões faciais e expressões espontâneas [16]. As emoções também são categorizadas em sete classes principais, as mesmas da FER2013. A grande diversidade das imagens faz da *RAF-DB* uma base valiosa para avaliar a capacidade dos modelos de aprendizado de máquina em lidar com a complexidade do reconhecimento emocional em condições não controladas.



Fig. 2. RAFDB. Fonte: <http://www.whdeng.cn/raf/model1.html>.

Por fim, a *KDEF* (*Karolinska Directed Emotional Faces*) é uma base de dados desenvolvida em ambiente controlado, composta por 4.900 imagens de 70 indivíduos, sendo 35 homens e 35 mulheres. Cada pessoa foi fotografada exibindo sete diferentes emoções (felicidade, tristeza, raiva, medo, surpresa, nojo e neutro) capturadas em cinco ângulos distintos (ver amostra na Figura 3). Por ser um conjunto de dados coletado em condições altamente controladas, a *KDEF* é ideal para treinar modelos que requerem consistência nas expressões faciais e uniformidade no ambiente [17, 18]. Sua organização sistemática contribui para um treinamento mais preciso e eficaz de redes neurais em tarefas de reconhecimento de emoções.

As bases *FER-2013*, *RAF-DB* e *KDEF* foram escolhidas devido à sua diversidade e complementaridade. A *FER-2013* oferece uma ampla gama de expressões faciais em várias

condições, enquanto a *RAF-DB* foca em emoções complexas. A *KDEF* é importante por conter imagens de diferentes ângulos, favorecendo a generalização do modelo. Juntas, elas fornecem um conjunto robusto para o reconhecimento de emoções em diferentes contextos e variações faciais.



Fig. 3. KDEF. Fonte: [https://www.researchgate.net/publication/331078578\\_Facial\\_emotion\\_detection\\_using\\_modified\\_eyemap-mouthmap\\_algorithm\\_on\\_an\\_enhanced\\_image\\_and\\_classification\\_with\\_tensorflow/figures?lo=1](https://www.researchgate.net/publication/331078578_Facial_emotion_detection_using_modified_eyemap-mouthmap_algorithm_on_an_enhanced_image_and_classification_with_tensorflow/figures?lo=1).

## V. PRÓXIMOS PASSOS

Com a definição do problema, dos objetivos e das bases de dados a serem utilizadas, os próximos passos desta pesquisa envolvem:

- Realizar a etapa de pré-processamento das imagens, incluindo redimensionamento, normalização e rotulação adequada das emoções-alvo.
- Aplicar técnicas de *data augmentation* para ampliar e balancear o conjunto de dados, aumentando a diversidade das expressões faciais.
- Implementar o modelo de rede neural utilizando a arquitetura ResNet50, explorando diferentes estratégias de *fine-tuning* para melhorar a performance na tarefa de reconhecimento emocional.
- Avaliar e comparar os resultados obtidos com diferentes configurações de hiperparâmetros e estratégias de treinamento.
- Investigar a possibilidade de integração do modelo em um sistema de monitoramento em tempo real, voltado à detecção de emoções em contextos de risco.

Essas etapas permitirão avançar para a construção de um sistema robusto de apoio à segurança da mulher, com foco em ambientes públicos.

O repositório com os arquivos adicionais estão disponíveis em: [https://github.com/MarianaLopeszzx/PCC146—Reconhecimento-de-Pradr-es\\_Mariana\\_Lopes](https://github.com/MarianaLopeszzx/PCC146—Reconhecimento-de-Pradr-es_Mariana_Lopes).

## REFERENCES

- [1] Instituto Patrícia Galvão, “Dossiê Feminicídio 2023,” Publicado online, 2023. Disponível em: <https://dossies.agenciapatriciagalvao.org.br/feminicidio/>
- [2] M. C. dos S. Silva, “Análise da violência de gênero no Brasil: um panorama atual,” *Revista de Políticas Públicas*, 2022. Disponível em: <https://periodicos.ufma.br/index.php/rppublica/article/view/20863>.
- [3] C. Oliveira e L. Silva, “Violência contra a mulher: impactos na saúde e políticas públicas de enfrentamento,” *Revista Brasileira de Saúde Materno Infantil*, 2021.

- [4] S. Russell e P. Norvig, "Artificial Intelligence: A Modern Approach," 3ª ed., Pearson, Boston, 2016.
- [5] I. Goodfellow, Y. Bengio e A. Courville, "Deep Learning," MIT Press, Cambridge, MA, 2016. Disponível em: <https://www.deeplearningbook.org/>.
- [6] P. Ekman, "Basic Emotions," in *Handbook of Cognition and Emotion*, T. Dalgleish e M. Power, Eds., John Wiley & Sons, 1999, pp. 45–60.
- [7] B. Li e D. Lima, "Facial expression recognition via ResNet-50," *International Journal of Cognitive Computing in Engineering*, vol. 2, art. 102248, 2021. DOI:10.1016/j.ijcce.2021.02.002.
- [8] R. K. Reghunathan, V. K. Ramankutty, A. Kallingal e V. Vinod, "Facial Expression Recognition Using Pre-trained Architectures," *Engineering Proceedings*, vol. 62, no. 1, art. 22, 2024. DOI:10.3390/engproc2024062022.
- [9] J. Kim, H. Lee e S. Park, "Performance Improvement of Speech Emotion Recognition Using ResNet Model with Data Augmentation–Saturation," *Applied Sciences*, vol. 15, no. 4, art. 2088, 2025. DOI:10.3390/app15042088.
- [10] S. Jayaraman e A. Mahendran, "CNN-LSTM based emotion recognition using Chebyshev moment and K-fold validation with multi-library SVM," *PLoS ONE*, vol. 20, no. 4, e0320058, Apr. 2025.
- [11] C. Białek, A. Mاتیolański e M. Grega, "An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks," *Electronics*, vol. 12, no. 2707, Jun. 2023.
- [12] E. S. Agung, A. P. Rifai e T. Wijayanto, "Image-based facial emotion recognition using convolutional neural network on Emognition dataset," *Sci. Rep.*, vol. 14, no. 14429, 2024.
- [13] A. K. Roy, H. K. Kathania, A. Sharma, A. Dey e M. S. A. Ansari, "ResEmoteNet: Bridging Accuracy and Loss Reduction in Facial Emotion Recognition," 2024.
- [14] B. Hasani, P. S. Negi e M. H. Mahoor, "BReG-NeXt: Facial Affect Computing Using Adaptive Residual Networks With Bounded Gradient," *IEEE Transactions on Affective Computing*, 2020.
- [15] Y. Khairuddin e Z. Chen, "Facial Emotion Recognition: State of the Art Performance on FER2013," Preprint, Boston University, May 2021. DOI:10.48550/arXiv.2105.03588.
- [16] X. Li, W. Deng, J. Du e outros, "Real-world Affective Faces Database (RAF-DB)," Introduzido em: Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild, 2020. Disponível em: <https://paperswithcode.com/dataset/raf-db>.
- [17] K. Goeleven, O. De Raedt, P. Leyman e R. Verschuere, "The Karolinska Directed Emotional Faces: a validation study," *Cognition and Emotion*, vol. 22, no. 6, pp. 1094–1118, 2008.
- [18] M. Garrido e C. Prada, "KDEF-PT: Valence, Emotional Intensity, Familiarity and Attractiveness Norms for 210 Faces in the Karolinska Directed Emotional Faces Database," *Frontiers in Psychology*, vol. 8, art. 2052, 2017.