



**CENTRO DE PESQUISA E DESENVOLVIMENTO TECNOLÓGICO EM INFORMÁTICA
E ELETROELETRÔNICA DE ILHÉUS- CEPEDI**

**ERICH BARRETO PEREIRA
MARIA FERNANDA CUNHA DA SILVA**

**Implementação e Análise de Classificação com Redes Convolucionais e o
dataset CUFS**

ILHÉUS – BAHIA

01/12/2024

RESUMO

Este projeto teve como objetivo desenvolver e avaliar um modelo de rede neural convolucional (CNN) para a classificação de imagens do dataset CUHK Face Sketch Database (CUFS), com a tarefa de categorizar faces com base no sexo biológico (masculino ou feminino). O processo foi estruturado em diversas etapas, incluindo a anotação manual das imagens, pré-processamento para padronização de dimensões e normalização, desenvolvimento de uma arquitetura personalizada de CNN e análise detalhada dos resultados. As avaliações do modelo foram conduzidas por meio de métricas como F1-Score, curva ROC e área sob a curva ROC (AUC-ROC), complementadas pela análise qualitativa das imagens classificadas incorretamente. O modelo apresentou um desempenho promissor, com alta precisão geral, mas revelou desafios específicos relacionados ao desbalanceamento das classes e às características das imagens classificadas incorretamente. Os resultados destacaram a importância do equilíbrio entre as classes e da qualidade do dataset na obtenção de classificações mais consistentes. Além disso, foi discutida a possibilidade de melhorias por meio de estratégias como aumento de dados, balanceamento do dataset e ajustes na arquitetura do modelo. A análise crítica dos resultados reforça o potencial das redes neurais convolucionais para a classificação de imagens, enquanto ressalta a necessidade de abordar limitações específicas do dataset e do modelo para alcançar uma maior generalização e desempenho.

Palavras-Chave: Rede Neural. CUHK. CUFS.

1 INTRODUÇÃO

1.1 Contextualização

O avanço da tecnologia na última década tem proporcionado uma explosão na criação e no uso de imagens digitais em diferentes domínios, desde redes sociais até aplicações médicas e de segurança. Essa crescente disponibilidade de dados visuais trouxe à tona a necessidade de desenvolver sistemas computacionais capazes de interpretar imagens de forma eficiente, seja para identificar objetos, reconhecer padrões ou categorizar elementos específicos. Dentro deste contexto, o problema de classificação de imagens se destaca como um dos mais desafiadores e importantes na área de visão computacional.

A classificação de imagens consiste na tarefa de atribuir um rótulo ou categoria predefinida a uma imagem com base em suas características visuais. Essa tarefa encontra aplicações em diversas áreas, como diagnóstico médico (detecção de doenças em radiografias), sistemas de segurança (reconhecimento facial), monitoramento ambiental (classificação de terrenos via satélite) e muitas outras. No entanto, a classificação eficiente e precisa de imagens requer técnicas avançadas capazes de capturar as complexidades inerentes às representações visuais, como variações de iluminação, ângulos de visualização, texturas e ruídos.

Tradicionalmente, métodos de classificação de imagens dependiam de técnicas baseadas em extração manual de características. Esses métodos exigiam que especialistas definissem previamente quais aspectos das imagens eram mais relevantes para a tarefa em questão, como bordas, cores e texturas. Apesar de úteis em alguns casos, essas abordagens apresentavam limitações importantes. A extração manual de características muitas vezes falhava em capturar informações mais profundas e complexas presentes em imagens, o que resultava em modelos menos robustos e com menor generalização em cenários reais.

Além disso, a enorme diversidade de variações nas imagens tornou inviável o uso de técnicas tradicionais para tarefas mais sofisticadas. Por exemplo, no reconhecimento facial, diferenças sutis na expressão, posição da câmera e iluminação podem alterar drasticamente a aparência de uma pessoa, dificultando a análise com métodos convencionais. Essa necessidade de compreender padrões complexos e de

alta dimensionalidade abriu espaço para o surgimento de métodos mais avançados, como as redes neurais convolucionais (CNNs).

1.2 Redes Convolucionais (CNNs)

Uma Rede Neural Convolucional (Convolutional Neural Network - CNN) é uma extensão das redes de Perceptrons Multicamadas, baseada no funcionamento do sistema visual humano. Similar aos métodos clássicos de visão computacional, uma CNN utiliza filtros para processar dados visuais, garantindo que as relações espaciais e de proximidade entre os pixels sejam mantidas ao longo das etapas de processamento na rede (JURASZEK, 2014).

As redes neurais convolucionais (RNCs) têm sido amplamente aplicadas em diversas áreas, como o processamento de imagens, vídeos, voz e áudio. No campo da classificação de imagens, elas desempenham um papel fundamental em tarefas como análise de imagens médicas, identificação de placas de trânsito, reconhecimento de objetos e interpretação de cenas. O Desafio de Reconhecimento Visual em Larga Escala do ImageNet (ILSVRC) tem sido uma força motriz para o avanço das RNCs, promovendo a avaliação contínua de diferentes arquiteturas utilizando uma base de dados extensa e diversificada com múltiplas categorias de objetos (Souza et al., 2020).

As redes neurais convolucionais (CNNs) possuem várias partes que desempenham funções específicas, sendo a camada de convolução uma das principais. Nessa etapa, neurônios aplicam filtros sobre regiões específicas da imagem, conectando-se apenas a subconjuntos de pixels da camada anterior. Cada conexão é associada a um peso, que, combinado com os dados de entrada, gera a saída transmitida para a próxima camada. Esses pesos podem ser representados como uma matriz, chamada de kernel, que define o filtro usado para processar a imagem (Vargas; Paes; Vasconcelos, 2016).

Ao contrário das redes totalmente conectadas, nas quais cada neurônio se conecta a todos os neurônios da camada anterior, as CNNs utilizam conexões locais, analisando apenas pequenas regiões da imagem. Neurônios que compartilham o mesmo filtro são organizados em mapas de características, e o compartilhamento de pesos entre esses neurônios reduz o número de parâmetros a serem ajustados e acelera o treinamento da rede (Vargas; Paes; Vasconcelos, 2016).

O tamanho do filtro define a área da imagem processada por cada neurônio, enquanto o passo (stride) determina o intervalo de pixels analisados, impactando o tamanho da saída gerada. Durante o treinamento, os filtros não são pré-definidos; em vez disso, a rede ajusta os pesos para identificar as características mais relevantes do conjunto de dados (Vargas; Paes; Vasconcelos, 2016).

Diferentemente das abordagens tradicionais de visão computacional, que exigem a definição manual de filtros, as CNNs aprendem automaticamente os padrões mais eficazes. Pesquisas mostram que as primeiras camadas de uma CNN geralmente detectam elementos simples, como bordas e cores, enquanto camadas mais profundas identificam características mais complexas. Essa hierarquia de extração de informações é uma das grandes vantagens das CNNs (Vargas; Paes; Vasconcelos, 2016).

Além disso, as CNNs podem criar filtros tridimensionais, analisando largura, altura e profundidade simultaneamente. Isso permite processar múltiplos canais de entrada, extraindo características mais complexas e detalhadas. Essa flexibilidade diferencia as CNNs das técnicas tradicionais, que geralmente se limitam a filtros bidimensionais, ampliando significativamente o potencial dessas redes para resolver problemas de análise de imagens (Vargas; Paes; Vasconcelos, 2016).

1.3 Problemática do Projeto

Neste projeto, o problema a ser abordado envolve a classificação de imagens de rostos humanos em duas categorias: masculino e feminino. O dataset CUHK Face Sketch Database (CUFS) fornece um conjunto de imagens que inicialmente foi desenvolvido para mapear fotografias para esboços artísticos. No entanto, para este trabalho, o objetivo será reinterpretar o dataset, focando na categorização das imagens com base no sexo biológico das pessoas retratadas.

Esse problema além de requerer um modelo capaz de diferenciar características faciais que identificam o sexo biológico, ele exige atenção especial na preparação dos dados, na escolha da arquitetura do modelo e na interpretação dos resultados. Variações nas condições das imagens, como iluminação e ângulos, podem influenciar o desempenho do modelo, destacando a importância de técnicas robustas de classificação, como as redes convolucionais.

Ao longo deste projeto, será explorado o papel das CNNs como uma solução moderna e eficaz para a classificação de imagens, com o objetivo de não apenas alcançar altos níveis de desempenho, mas também compreender os desafios e as limitações associados ao treinamento de modelos convolucionais a partir de dados brutos.

2 METODOLOGIA

A metodologia do projeto foi dividida em várias etapas desde a coleta das imagens até a análise dos resultados obtidos da rede neural. A primeira etapa foi a coleta das imagens, rotulação e análise das imagens. Foi necessário a rotulação inicial das imagens, sendo 0 para homens e 1 para mulher. Para o armazenamento destas rotulações, foi criado um arquivo no formato .csv via código, que apresentava cada imagem e solicita ao usuário que insira o número referente ao sexo da imagem.

Esta análise inicial e rotulagem dos dados considerou as semelhanças entre indivíduos do mesmo sexo e características extras como a presença de barba, bigode ou adereços no cabelo. Também foi analisado especificidades de algumas imagens, como formatos diferentes dos rostos, boca, como também mulheres de cabelo curto e homens que cabelo mais longo. Estas imagens intermediárias podem gerar erros de previsões ou diminuir a acurácia do treinamento. Ao final, foram indicadas 54 fotografias sendo do sexo feminino e 134 imagens do sexo masculino. Foi deixado ao usuário a escolha de utilizar a análise feita pelos desenvolvedores, ou gerar a própria rotulação, a fim de olhar e corrigir possíveis incongruências.

Após a rotulagem e análise dos dados, gerou-se a divisão das imagens para treinamento, validação e testes da rede neural. Conforme pedido em atividade, treinamento ficou com 50%, validação com 30% e testes com 20% utilizando a metodologia de separação $seed = 23$. Além disso, redimensionou-se todas as imagens para o tamanho padrão de 250x200 pixels e foi normalizado para que tivesse valores entre 0 e 1.

Para o treinamento da rede neural foram utilizadas as seguintes configurações: a primeira camada composta por uma convolução 2D com 32 filtros com tamanho 3x3 pixels; a função de ativação relu, saída unidimensional com a função flattlen, camada densa com 128 neurônios e função de ativação relu, e a camada de saída com os 2 neurônios de saída para a classificação entre masculino e feminino. As escolhas dos valores de filtro e camada densa foram de acordos com a melhor eficiência do projeto durante os testes e validações, utilizando diferentes divisões de dados.

No treinamento, foram indicados como otimizador “Adam”, no qual é um algoritmo baseado em gradiente de funções objetivas com base na estimativa adaptada e com o fator de aprendizado em 0,0001. A escolha deste valor ocorreu por conta da distribuição irregular entre as fotográficas de homem e mulher, trazendo

insegurança na identificação das fotos com sexo feminino. Além disso, o sistema considerou como cálculo de perdas a discrepância entre as distribuições preditas pelo modelo e as distribuições reais das 2 variáveis, utilizando o *binary_crossentropy*; e para métricas de regulação foi utilizado a acurácia.

Na validação dos dados, foi considerado a utilização de 30% dos dados. Utilizando o método *evaluate*, a sua saída indicou o valor de perdas do sistema aos dados de validação e a acurácia do sistema.

Em seguida, foi utilizado a CNN para realizar testes, avaliando a generalização, utilizando as métricas de acurácia, precisão, *recall*, F1, e criado a curva ROC. Além disso, foram apresentadas as imagens que tiveram falsa previsão, para que o usuário possa avaliar características em que o sistema apresentou valor divergente do real.

3 DISCUSSÃO

O desempenho do modelo de rede neural convolucional (CNN) foi avaliado a partir das métricas de precisão, recall, F1-Score, acurácia, e análise da curva ROC, além da inspeção das imagens classificadas incorretamente. Estas métricas fornecem uma visão sobre o comportamento do modelo, permitindo uma análise sobre a eficácia da arquitetura e das decisões tomadas durante o treinamento. A partir dos resultados apresentados, foi possível explorar diversos aspectos que ajudam a entender tanto os pontos fortes quanto as limitações do modelo.

O F1-Score geral de 0.89 é um indicador de que o modelo apresenta um bom equilíbrio entre precisão e recall, especialmente considerando o desbalanceamento entre as classes. O F1-Score, oferece uma medida robusta quando é apresentado uma distribuição desigual entre as classes, como foi o caso deste modelo, onde a quantidade de imagens de homens foi significativamente maior do que a de mulheres. Com um resultado de 0.89, o modelo consegue balancear bem o risco de falsos positivos e falsos negativos, o que é essencial em problemas de classificação binária.

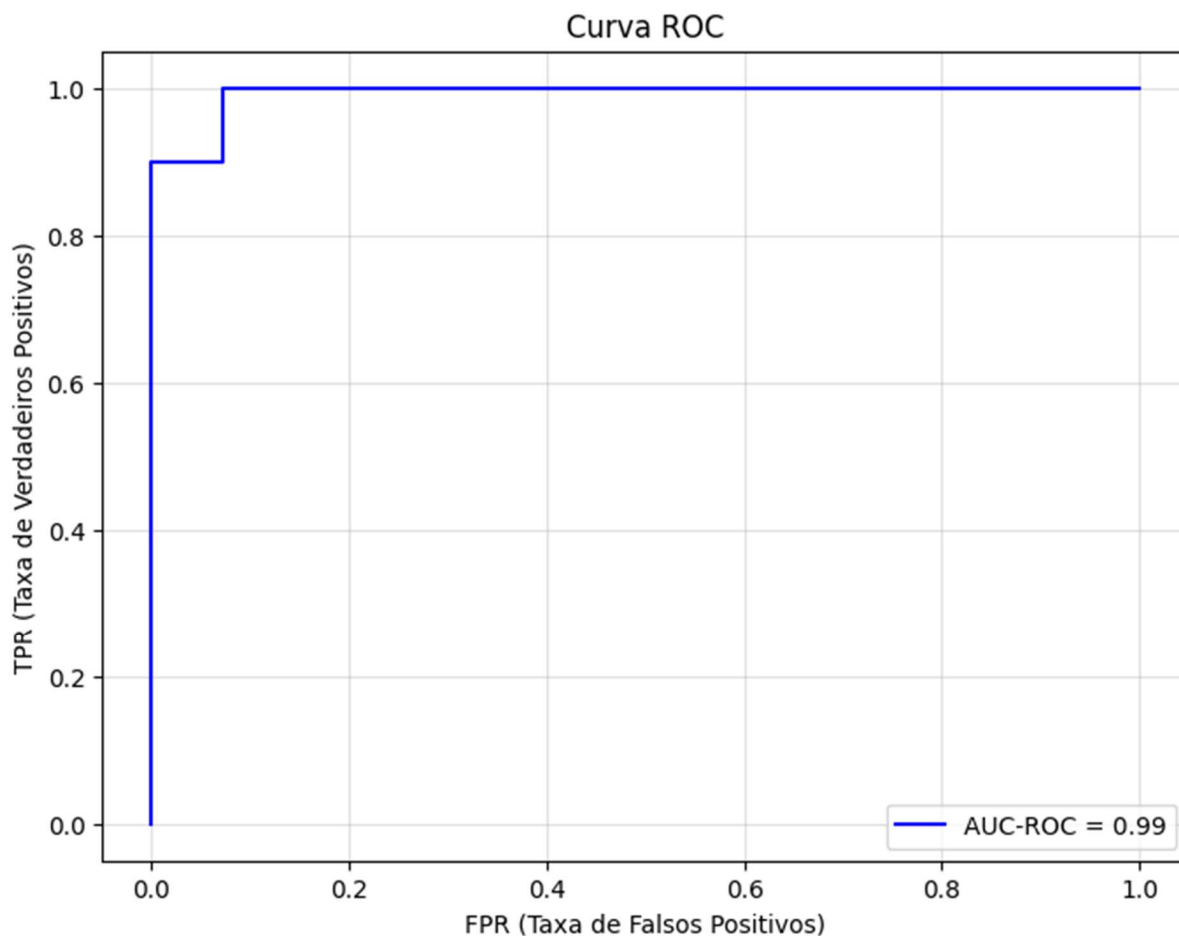
No entanto, após ter sido feito uma análise mais detalhada das métricas específicas para cada classe, foi revelado áreas que podem ser melhoradas. Para a classe "masculino" (0), a precisão foi de 0.93, e o recall foi de 0.96, resultando em um F1-Score de 0.94. Os resultados obtidos são bastante satisfatórios e indicam que o modelo tem um bom desempenho ao identificar imagens de homens, com um baixo número de falsos negativos. Por outro lado, a classe "feminino" (1) apresentou um

desempenho diferente, com precisão de 1.00 e recall de 0.80, resultando em um F1-Score de 0.89. A precisão perfeita na classe feminina sugere que, quando o modelo classifica uma imagem como feminina, ele está quase sempre correto. No entanto, o recall mais baixo indica que o modelo deixou de classificar corretamente algumas imagens femininas, o que pode ser um reflexo das dificuldades associadas à representatividade da classe feminina no dataset, já que a quantidade de imagens de mulheres foi significativamente menor.

A diferença nos resultados entre as duas classes pode indicar que o modelo tem maior dificuldade em identificar corretamente as imagens de mulheres. Isso pode ser atribuído ao fato de que, como foi mencionado na metodologia, o dataset apresenta um desbalanceamento entre as classes, isso pode afetar a capacidade do modelo de aprender as características associadas ao sexo feminino. Além disso, como o número de imagens de mulheres é menor, o modelo pode ter visto menos exemplos durante o treinamento, o que pode ter gerado uma performance inferior nessa classe. Uma solução possível para mitigar esse problema seria aplicar técnicas de balanceamento, como o *oversampling* da classe minoritária ou o uso de pesos ajustados para as classes durante o treinamento.

Além das métricas, na Figura 1 também foi feita a análise da curva ROC e da área sob a curva (AUC).

Figura 1 – Análise da Curva ROC e da área sob a curva

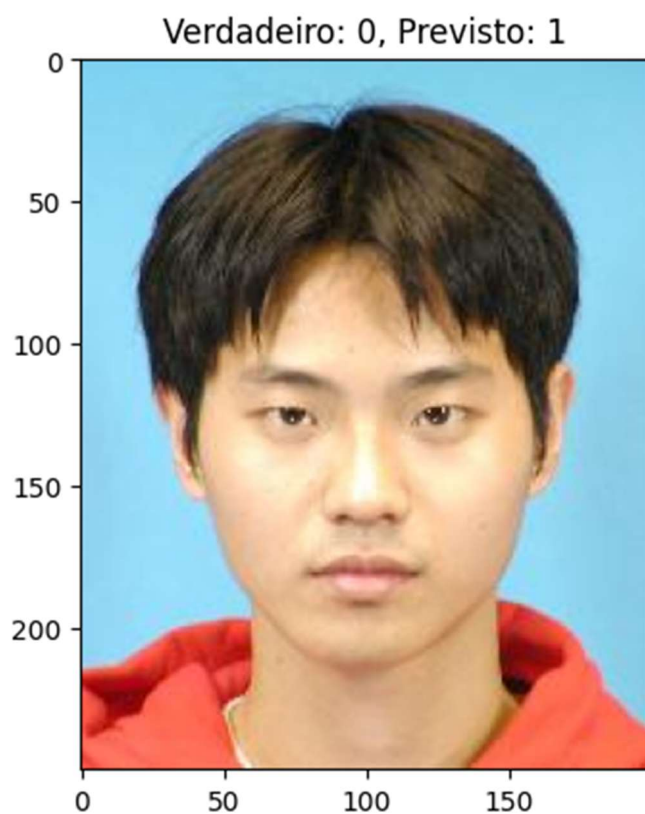


Fonte: Elaboração própria.

A curva ROC revelou que o modelo tem uma capacidade razoável de discriminar entre as duas classes, o que é confirmado pela AUC de 0.95. A AUC próximo de 1.00 indica que o modelo tem uma excelente capacidade de separação entre as classes, o que é um bom sinal de que a rede neural aprendeu a distinguir as características essenciais para a classificação correta das imagens. A curva ROC também revelou que o modelo, embora eficiente na maioria das classificações, tem uma leve inclinação em favor da classe masculina, o que reforça a necessidade de ajustes, especialmente em relação ao balanceamento de classes.

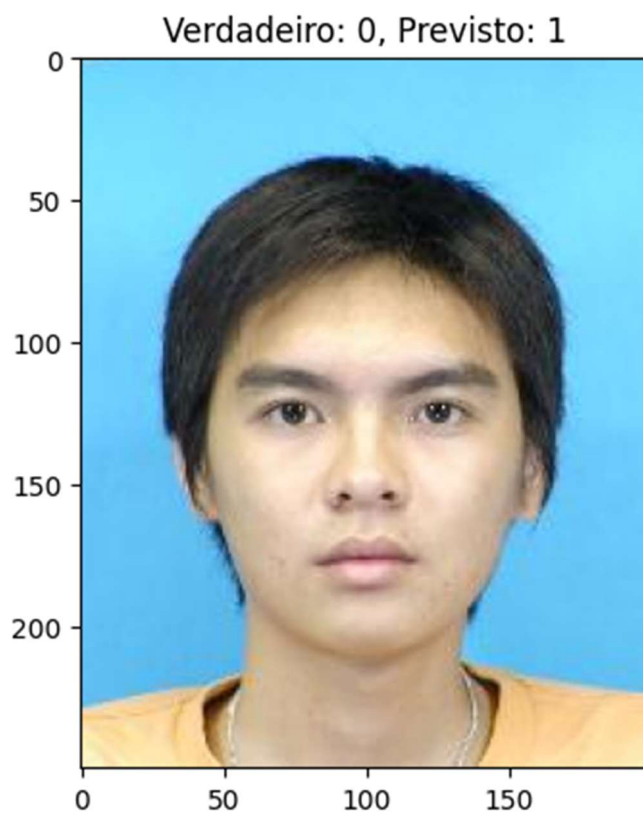
Em relação à análise das imagens classificadas incorretamente, foi possível observar que as falhas do modelo ocorreram com imagens que apresentam características faciais mais ambíguas ou que não seguem os padrões típicos de cada sexo. No modelo, foram identificadas 3 imagens que foram classificadas incorretamente:

Figura 2 – Imagem classificada incorretamente pelo modelo



Fonte: CUHK Face Sketch Database (CUFS)

Figura 3 – Imagem classificada incorretamente pelo modelo



Fonte: CUHK Face Sketch Database (CUFS)

Figura 4 – Imagem classificada incorretamente pelo modelo



Fonte: CUHK Face Sketch Database (CUFS)

Algumas imagens de mulheres com cabelo curto foram classificadas de forma errada, o que sugere que o modelo teve dificuldade em lidar com essas variações no formato dos rostos e nos estilos de cabelo. Além do mais, imagens com iluminação ruim ou ângulos incomuns também podem ter contribuído para obter classificações incorretas.

Esses tipos de variação são comuns em problemas de visão computacional e refletem a necessidade de maior robustez do modelo em relação a condições variadas de captura de imagem. O modelo pode ser aprimorado através de técnicas de aumento de dados (*data augmentation*), como rotação, escala e alteração de brilho, para tornar o modelo mais robusto a essas variações.

Outro ponto importante a ser considerado é a limitação do dataset. Embora o dataset CUHK Face Sketch Database (CUFS) seja um bom ponto de partida para tarefas de classificação de faces, ele apresenta algumas limitações que podem ter influenciado os resultados. O desbalanceamento entre as classes, com mais imagens de homens do que de mulheres, pode ter afetado a capacidade do modelo em aprender características representativas da classe feminina. Além disso, a qualidade das imagens, em termos de resolução e iluminação, não é sempre consistente, o que pode ter impactado a capacidade do modelo de extrair características precisas de cada imagem. Uma solução seria buscar datasets mais balanceados ou aplicar técnicas de *data augmentation* para melhorar a diversidade dos dados.

Para melhorar o desempenho do modelo, algumas modificações na arquitetura da rede neural podem ser úteis. Uma possibilidade seria o aumento da profundidade da rede, adicionando mais camadas convolucionais ou camadas de pooling, isso poderia ajudar o modelo a capturar características mais complexas das imagens. Além disso, ajustes nos hiperparâmetros, como o aumento do número de neurônios na camada densa ou a utilização de uma função de ativação diferente, como o *Leaky ReLU*, poderiam melhorar a capacidade do modelo de lidar com dados desbalanceados e com as variações de características faciais. Outra estratégia seria a aplicação de técnicas de regularização, como *dropout*, para evitar o overfitting, especialmente considerando que o modelo foi treinado com um número relativamente pequeno de imagens.

Por fim, uma surpresa durante a execução do projeto foi a diferença de desempenho entre as duas classes, com o modelo apresentando um desempenho significativamente melhor na classificação das imagens masculinas. Esse resultado evidenciou a importância de um maior equilíbrio nas classes, não apenas no número de imagens, mas também na diversidade e qualidade das imagens de cada classe. Embora a rede neural tenha alcançado bons resultados gerais, é evidente que o modelo ainda precisa de ajustes para lidar melhor com a classe feminina, uma vez que a precisão nessa classe foi muito alta, mas o recall foi consideravelmente mais baixo.

4 CONCLUSÃO

Este projeto demonstrou a eficácia de uma rede neural convolucional (CNN) aplicada à classificação de imagens do dataset CUHK Face Sketch Database (CUFS), com o objetivo de distinguir entre imagens masculinas e femininas. Ao longo do desenvolvimento, foi possível observar que a metodologia adotada, que envolveu a coleta, rotulação e pré-processamento cuidadoso dos dados, contribuiu significativamente para a construção do modelo. A arquitetura da CNN, que inclui camadas de convolução, camadas densas e funções de ativação apropriadas, foi bem-sucedida para o problema de classificação binária. A avaliação do modelo, através de métricas como F1-Score, AUC-ROC e precisão, forneceu uma visão clara da performance, destacando a importância de otimizar o treinamento, principalmente considerando o desbalanceamento entre as classes.

No entanto, algumas limitações foram identificadas, especialmente relacionadas ao desbalanceamento entre as imagens de homens e mulheres no dataset. A quantidade de fotos de mulheres foi significativamente menor do que a de homens, o que pode ter prejudicado a performance do modelo na classificação feminina. Além disso, algumas imagens com características atípicas apresentaram resultados menos precisos, indicando que o modelo pode ter dificuldades em generalizar para tais variações.

Para trabalhos futuros, sugerem-se algumas melhorias que poderiam aumentar a eficácia do modelo. Primeiramente, seria ideal aumentar o número de imagens de mulheres no dataset, a fim de equilibrar as classes e melhorar o desempenho do modelo nessa categoria. Além disso, a inclusão de um maior número de imagens, no geral, permitiria ao modelo aprender com mais exemplos, tornando-o mais robusto e capaz de lidar com uma variedade maior de situações. Outra recomendação é realizar uma análise mais aprofundada das imagens classificadas incorretamente para entender melhor as causas dos erros, e aplicar técnicas como *data augmentation* para lidar com as variações nas imagens (por exemplo, mudanças de iluminação, ângulos ou distorções faciais). A utilização de métodos para corrigir o desbalanceamento de classes, como o ajuste de pesos ou *oversampling* nas classes menos representadas, também poderia ser considerada para melhorar ainda mais os resultados.

REFERÊNCIAS

VARGAS, Ana Caroline Gomes; PAES, Aline; VASCONCELOS, Cristina Nader. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: **Proceedings of the xxix conference on graphics, patterns and images**. sn, 2016. Disponível em: <http://gibis.unifesp.br/sibgrapi16/e-proceedings/wuw/7.pdf> Acesso em: 30 nov. 2024.

JURASZEK, Guilherme Defreitas et al. Reconhecimento de produtos por imagem utilizando palavras visuais e redes neurais convolucionais. 2014. Disponível em: https://www.researchgate.net/profile/Guilherme-Juraszek-2/publication/275958578_Image_Recognition_of_Products_Using_Bag_of_Visual_Words_and_Convolutional_Neural_Networks/links/554b6b7e0cf29752ee7c7cda/Image-Recognition-of-Products-Using-Bag-of-Visual-Words-and-Convolutional-Neural-Networks.pdf. Acesso em: 30 nov. 2024.

SOUZA, Victor et al. Análise comparativa de redes neurais convolucionais no reconhecimento de cenas. **Anais do Computer on the Beach**, v. 11, p. 419-426, 2020. Disponível em: <https://periodicos.univali.br/index.php/acotb/article/view/16801>. Acesso em: 30 nov. 2024.

FERREIRA, Alessandro dos Santos. Redes neurais convolucionais profundas na detecção de plantas daninhas em lavoura de soja. 2017. Disponível em: <https://repositorio.ufms.br/handle/123456789/3101>. Acesso em: 30 nov. 2024.

OLIVEIRA, Petronio Diego Silva de. Uso de aprendizagem de máquina e redes neurais convolucionais profundas para a classificação de áreas queimadas em imagens de alta resolução espacial. 2020. Disponível em: <http://www.realp.unb.br/jspui/handle/10482/38234>. Acesso em: 30 nov. 2024.