

**Relatório Técnico: Implementação e Análise de Classificação  
com Redes Convolucionais e o dataset CUFS  
Lucas Silva de Oliveira e Samy Celes Barrozo  
17/11/2024**

## Resumo

O objetivo deste trabalho foi utilizar o dataset CUHK Face Sketch Database (CUFS) para classificar imagens de rostos em duas categorias: masculino e feminino, com base no sexo biológico. Para isso, empregamos redes neurais convolucionais (CNNs), um tipo de modelo amplamente utilizado para tarefas de visão computacional devido à sua capacidade de capturar padrões espaciais em imagens.

O processo envolveu um pré-processamento dos dados, com redimensionamento, normalização e balanceamento dos dados via oversampling para tratar a desproporção entre as classes, uma separação em conjuntos de treino, validação e teste (50%, 30% e 20% dos dados restantes, respectivamente), a criação de uma CNN com camadas convolucionais, de pooling, dropout, e densas, usando a função de ativação ReLU e a função de saída sigmoid e, por fim, o modelo foi treinado por 10 épocas com otimizador Adam e função de perda `binary_crossentropy`. Avaliamos o modelo utilizando acurácia e perda nos conjuntos de validação e teste.

Foi observado que o modelo alcançou uma acurácia de 92.59% no conjunto de teste e apresentou uma perda de 0.257, indicando um bom desempenho na tarefa de classificação. As curvas de acurácia e perda mostraram um treinamento estável, sem indícios significativos de overfitting ou underfitting, resultando em um modelo confiável e bem generalizado.

## Introdução

A classificação de imagens é um problema fundamental em visão computacional, com aplicações em diversas áreas, como segurança, diagnóstico médico e reconhecimento facial. O objetivo desses problemas é associar uma ou mais categorias a uma imagem com base nas suas características visuais.

Nesse contexto, as Redes Neurais Convolucionais (CNNs) revolucionaram o campo da visão computacional, superando algoritmos tradicionais em tarefas como classificação de imagens, detecção de objetos e segmentação. O principal diferencial das CNNs é sua capacidade de aprender representações hierárquicas de imagens, identificando características

simples, como bordas e texturas, em camadas iniciais, e características mais complexas, como formas e objetos, em camadas mais profundas, em um padrão de arquitetura que se assemelha a organização do córtex visual.

No contexto deste trabalho, exploramos o potencial das CNNs para classificar imagens do dataset CUHK Face Sketch Database (CUFS), originalmente projetado para mapear fotografias de rostos para desenhos artísticos. Aqui, o desafio foi diferente: determinar o sexo biológico dos indivíduos representados nas imagens. Esta tarefa é desafiadora devido à sutileza das diferenças visuais, muitas vezes presentes entre os rostos masculinos e femininos, exigindo que o modelo seja capaz de capturar padrões precisos e generalizáveis.

As redes convolucionais, com sua habilidade de identificar e aprender essas características, tornam-se a abordagem ideal para este problema, sendo um exemplo claro de como avanços em aprendizado profundo podem ser aplicados para resolver problemas de classificação de imagens de maneira eficiente.

## **Metodologia**

O processo de desenvolvimento deste modelo de classificação de imagens começou com a preparação dos dados, utilizando o CUHK Face Sketch Database (CUFS), um dataset originalmente criado para mapear fotografias de rostos para desenhos artísticos. Para esta tarefa, o objetivo foi classificar as imagens de faces de acordo com o sexo biológico, ou seja, identificar se a face é masculina ou feminina.

A primeira etapa envolveu o carregamento das imagens e a criação de rótulos, que indicam o sexo biológico das pessoas nas imagens. Para garantir que o modelo pudesse aprender de maneira eficaz, foi necessário realizar um pré-processamento nas imagens. As imagens foram redimensionadas para um tamanho uniforme de 250x200 pixels. Este redimensionamento foi feito para padronizar o tamanho das entradas, já que imagens de diferentes tamanhos podem dificultar o treinamento. Além disso, as imagens foram normalizadas, ou seja, os valores dos pixels foram escalados para o intervalo  $[0, 1]$  ao dividir os valores originais por 255, o que facilita a convergência do modelo durante o treinamento, pois valores elevados podem dificultar o processo de ajuste dos pesos.

O próximo passo foi dividir o conjunto de dados em três partes: treinamento, validação e teste. O conjunto de treinamento foi responsável por ensinar o modelo, utilizando 50% do total de imagens. A validação, com 30% dos dados, foi usada para ajustar os hiperparâmetros e monitorar o desempenho do modelo durante o treinamento. O restante (20%) foi reservado para o conjunto de teste, que seria utilizado somente após o treinamento para avaliar o desempenho final do modelo.

A arquitetura do modelo proposto é uma rede neural convolucional (CNN), que é comumente utilizada em problemas de classificação de imagens devido à sua capacidade de extrair características espaciais das imagens. O modelo foi projetado com três camadas convolucionais, cada uma seguida de uma camada de MaxPooling. A primeira camada convolucional utiliza 32 filtros de tamanho 3x3, a segunda camada aumenta para 64 filtros, e a terceira para 128 filtros. O MaxPooling foi utilizado para reduzir a dimensionalidade das imagens e diminuir a complexidade computacional, mantendo as características mais importantes. Entre as camadas convolucionais, foram aplicados dropouts para evitar o overfitting, com taxas de 25% após a primeira camada e 50% antes da camada densa.

Dropout é uma técnica de regularização que consiste em desativar aleatoriamente uma porcentagem específica de neurônios em uma camada durante o treinamento do modelo. Essa abordagem força a rede neural a não depender de nenhum neurônio específico, favorecendo a aprendizagem de um conjunto mais diversificado de características e promovendo uma maior capacidade de generalização.

Após a extração das características, a saída das camadas convolucionais foi achatada utilizando a camada Flatten, e em seguida foi conectada a uma camada densa com 128 neurônios e ativação ReLU. A última camada foi uma camada densa com um único neurônio e ativação sigmoid, ideal para problemas de classificação binária, como é o caso da nossa tarefa (masculino ou feminino).

O modelo foi compilado utilizando o otimizador Adam, que é eficiente na atualização dos pesos do modelo, combinando as vantagens de outros algoritmos como o SGD e RMSProp. A função de perda utilizada foi a `binary_crossentropy`, apropriada para

classificação binária, e a métrica de avaliação foi a acurácia, para acompanhar o desempenho do modelo durante o treinamento.

Após a definição da arquitetura, o modelo foi treinado utilizando o conjunto de dados de treinamento, por 10 épocas, com um tamanho de lote (batch size) de 32. Durante o treinamento, também foram avaliados os dados de validação, a fim de verificar a generalização do modelo e ajustar os parâmetros, caso necessário.

Para avaliar o desempenho do modelo, após o treinamento, utilizamos o conjunto de testes, com o objetivo de verificar como o modelo se comporta em dados não vistos durante o treinamento. Além da acurácia tradicional, outras métricas foram calculadas, como o F1-score e o AUC-ROC, para garantir uma análise mais completa. O F1-score é particularmente útil quando se trata de avaliar o equilíbrio entre precisão e recall, especialmente em problemas desbalanceados, enquanto o AUC-ROC fornece uma visão geral da capacidade do modelo em distinguir entre as classes, independente do limiar de decisão.

Esse conjunto de avaliações permitiu uma análise detalhada do desempenho do modelo, destacando suas forças e áreas que poderiam ser aprimoradas.

## Discussão

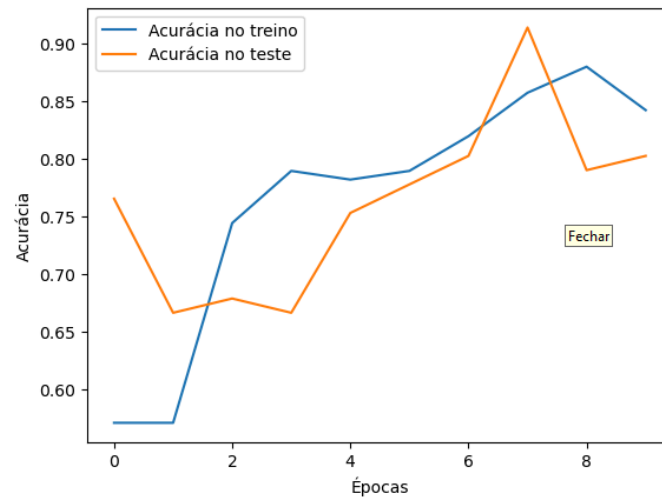
O modelo atingiu F1-score de 0,90 indicando que o mesmo conseguiu alcançar um bom equilíbrio entre as métricas de *precision* e *recall*, evidenciando que ele não apenas classifica bem as imagens positivas, mas também minimiza os falsos positivos e falsos negativos. Este resultado, evidenciado nos gráficos abaixo, é especialmente importante em um problema de classificação binária, onde manter o equilíbrio entre as classes é essencial para a generalização do modelo.

**Figura 1** - Métricas utilizadas e Valores Atingidos

Métrica	Valor
Loss (Perda)	0.256900
Acurácia	0.925926
F1-Score	0.904762
AUC	0.982955

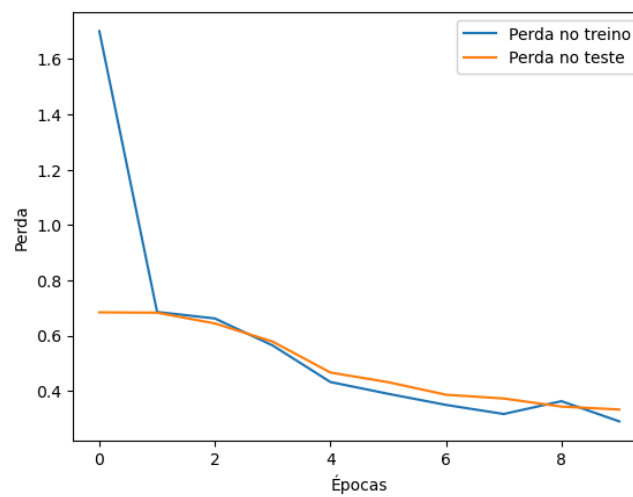
FONTE: Autores

**Figura 2 - Gráfico de Acurácia do Modelo**



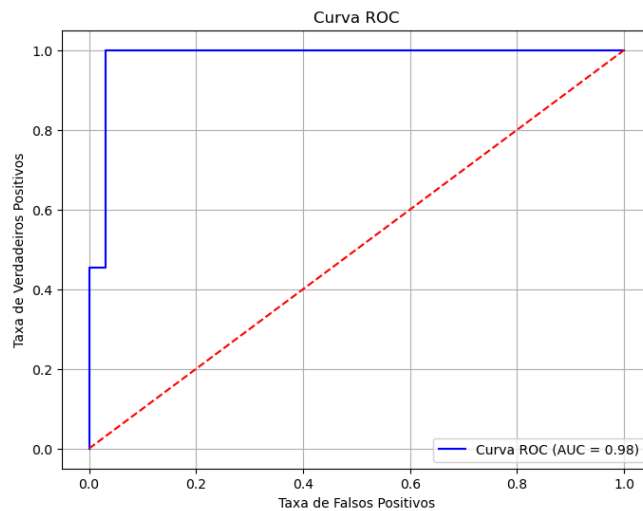
FONTE: Autores

**Figura 3 - Gráfico de Perda do Modelo**



FONTE: Autores

**Figura 4 - Gráfico da Curva ROC**



FONTE: Autores

No entanto, a análise qualitativa das imagens classificadas incorretamente sugere que o modelo enfrentou dificuldades específicas com certos tipos de imagens. Por exemplo, características ambíguas, como homens com rostos mais delicados e cabelos longos ou mulheres com traços mais marcantes, como bochechas pronunciadas ou olhos bem definidos, parecem ter confundido o modelo. Essas características fogem aos padrões faciais de gênero que o modelo pode ter aprendido, indicando que ele está sensível a aspectos faciais menos padronizados.

Uma possível limitação do dataset que pode ter influenciado os resultados é a qualidade e a diversidade das imagens. Apesar de o dataset ter sido balanceado entre as classes, a repetição de imagens pode ter reduzido a variabilidade e dificultado o aprendizado de padrões mais robustos. Além disso, se as imagens apresentarem ruído, variações de iluminação ou ângulos atípicos, isso também pode ter impactado a performance do modelo. Essas limitações reforçam a importância de utilizar datasets diversificados e bem representados.

Para melhorar o desempenho, algumas modificações na arquitetura do modelo ou nos hiperparâmetros poderiam ser exploradas, como testar taxas de *learning rate* menores ou usar técnicas como *learning rate scheduler* para otimizar o aprendizado, implementar transformações como rotação, alterações de brilho e contraste para criar mais variações nas

imagens; experimentar outras técnicas de regularização, como *weight decay* ou camadas de *dropout* mais ajustadas; utilizar modelos pré-treinados como base, como ResNet ou VGG, para aproveitar aprendizado transferido de tarefas semelhantes.

Por fim, uma descoberta interessante durante a realização da atividade foi o impacto das características ambíguas no desempenho do modelo. Embora ele tenha apresentado resultados satisfatórios em termos de métricas, a análise qualitativa revelou que o modelo ainda encontra desafios em capturar nuances mais sutis nas imagens faciais, como exemplificado nas imagens abaixo:

**Figura 5 - Faces classificadas incorretamente**



FONTE: Autores

Além disso, a constatação de que redes convolucionais profundas, mesmo em datasets relativamente pequenos, demandam considerável tempo de execução foi um aprendizado valioso, destacando a importância de se considerar a eficiência computacional no design do modelo.



## Conclusão e Trabalhos Futuros

O projeto demonstrou com sucesso a aplicação de redes neurais convolucionais para classificar imagens de faces humanas por sexo biológico, tendo como fonte de dados o dataset "CUHK Face Sketch Database (CUFS)". O modelo alcançou métricas sólidas, como F1-Score de 0.90 e uma AUC de 0.98, indicando que foi capaz de equilibrar com eficácia precisão e recall, além de separar eficientemente as classes. Esses resultados demonstram que a arquitetura proposta é eficaz e pode ser generalizada para problemas de cunho semelhante. Ademais, um ponto importante a se observar é que a análise qualitativa das previsões incorretas também revelou que características faciais ambíguas podem ser um fator que dificulta a classificação.

No entanto, mesmo com métricas positivas, alguns desafios foram identificados. O tempo de execução, mesmo com um dataset pequeno, foi relativamente alto devido à complexidade das operações convolucionais. Além disso, algumas limitações do dataset, como o uso de imagens estilizadas e um possível desbalanceamento implícito em representações faciais, podem ter influenciado os resultados.

Tendo em vista pontos de melhora em trabalhos futuros, alguns pontos de aprimoramento surgem como : procurar dataset com uma maior diversidade das imagens, incluindo variações de iluminação, ângulos e expressões faciais; garantir que o dataset seja balanceado não apenas em termos de classes, mas também em características faciais para evitar viés. Com relação a arquitetura do modelo, seria interessante experimentar mais livremente diferentes tamanhos de filtros e profundidades para ajustar o modelo à natureza estilizada das imagens.

Outras melhorias possíveis incluem realizar busca em grade para identificar combinações ótimas de hiperparâmetros, como taxa de aprendizado e tamanho de lote; implementar técnicas como *dropout* ou *batch normalization* para melhorar a generalização. Aprofundar os estudos para aplicação de técnicas de explicabilidade de modelos, como *Grad-CAM*, para identificar quais regiões das imagens influenciam as decisões do modelo e utilizar análises pós-modelo para identificar padrões em imagens classificadas incorretamente.

Por fim, explorar o uso de transfer learning para adaptar o modelo a outros tipos de classificação, como idade ou expressões faciais e estender o projeto para realizar tarefas como reconstrução de imagens ou segmentação facial, aprofundando o uso de redes convolucionais são também pontos importantes de melhoria que podem ser foco de estudo e aplicação no desenvolvimento futuro do projeto.

## Referências

ALVES, Gisely. **Entendendo Redes Convolucionais (CNNs)**. Medium, 2018. Disponível em: <https://medium.com/neuronio-br/entendendo-redes-convolucionais-cnns-d10359f21184>. Acesso em: 1 dez. 2024.

**Camada convolucional**. Databricks, 2024. Disponível em: <https://www.databricks.com/br/glossary/convolutional-layer>. Acesso em: 1 dez.2024.

CLAPPIS, Alan Melo. **Uma introdução às redes neurais convolucionais utilizando o Keras**. Medium, 2019. Disponível em: <https://medium.com/data-hackers/uma-introdu%C3%A7%C3%A3o-as-redes-neurais-convolucionais-utilizando-o-keras-41ee8dcc033e>. Acesso em: 1 dez. 2024.

CHUGH, Vidhi. **AUC e a curva ROC no aprendizado de máquina**. Datacamp, 1 out. 2024. Disponível em: <https://www.datacamp.com/pt/tutorial/auc>. Acesso em: 1 dez. 2024.

ESTEVES, Toni. **Uma introdução aos diferentes tipos de convoluções**. Medium, 2022. Disponível em: <https://estevestoni.medium.com/uma-introdu%C3%A7%C3%A3o-aos-diferentes-tipos-de-convolu%C3%A7%C3%B5es-d3ce7cd81759>. Acesso em: 1 dez. 2024.

**F-1 SCORE**. Scikit-learn, 2024. Disponível em: [https://scikit-learn.org/1.5/modules/generated/sklearn.metrics.f1\\_score.html](https://scikit-learn.org/1.5/modules/generated/sklearn.metrics.f1_score.html). Acesso em: 1 dez. 2024.

FILHO, Mário. **Precisão, Recall e F1 Score em Machine Learning**. MarioFilho, 2023. Disponível em: <https://mariofilho.com/precisao-recall-e-f1-score-em-machine-learning/>. Acesso em: 1 dez. 2024.

**O que são redes neurais convolucionais?**.IBM, 2024. Disponível em: <https://www.ibm.com/br-pt/topics/convolutional-neural-networks>. Acesso em: 01 de dez. 2024.

PÁDUA, Mateus. **Machine Learning - Métricas de avaliação: Acurácia, Precisão e Recall, F1-score**. Medium, 2020. Disponível em:  
<https://medium.com/@mateuspdua/machine-learning-m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-e-recall-d44c72307959>. Acesso em 1 dez. 2024.