Classificação de imagens usando Redes Neurais Convolucionais

Projeto final da disciplina SIN 393 - Visão Computacional

1st Felipe Pereira Rodrigues
Matrícula: 7610

2nd Herbert Ribeiro Sampaio
Matrícula: 7633

3rd Luiz Davi Vieira Alves
Matrícula: 6335

***Abstract*—Este trabalho apresenta um estudo abrangente sobre classificação de imagens utilizando Redes Neurais Convolucionais (CNNs). O projeto explora o uso de modelos predefinidos do PyTorch, incluindo a AlexNet e a ResNet-50, e compara seus resultados com uma CNN personalizada. A análise inclui experimentos detalhados no conjunto de dados Animals10, abordando métricas como Acurácia, Precisão, Recall e F1-Score, além de matrizes de confusão. Gráficos e tabelas são apresentados para ilustrar os resultados e demonstrar a robustez dos modelos avaliados.**

***Index Terms*—classificação de imagens, redes neurais convolucionais, aprendizado profundo, PyTorch, visão computacional**

I. Introdução

O campo de aprendizado profundo tem avançado significativamente nos últimos anos, especialmente em visão computacional, onde Redes Neurais Convolucionais (CNNs) se destacam por sua capacidade de capturar padrões em imagens, viabilizando soluções para classificação, detecção e segmentação de objetos. Essas redes revolucionaram setores como saúde, segurança, transporte e entretenimento.

A classificação de imagens tem se mostrado essencial em tarefas do mundo real, como diagnóstico médico automatizado e reconhecimento facial. Implementar CNNs requer compreender desde o processamento dos dados até o ajuste de hiperparâmetros e avaliação de modelos. Este estudo compara arquiteturas modernas, como a ResNet-50, com uma CNN personalizada no conjunto de dados Animals10, que representa um desafio devido à variedade de iluminação, posição e contexto das imagens.

O principal objetivo é demonstrar as vantagens e limitações de diferentes arquiteturas, destacando análises de métricas e matrizes de confusão, e fornecendo insights para o avanço no uso de CNNs em tarefas complexas de classificação.

Por meio deste estudo, avaliamos o desempenho das arquiteturas AlexNet e ResNet-50 em um dataset desafiador, analisando suas métricas de desempenho e propondo direções para estudos futuros. As contribuições incluem uma análise detalhada das métricas, visualizações de erros nas matrizes de confusão e insights sobre como arquiteturas modernas impactam positivamente o desempenho em tarefas complexas de visão computacional.

II. Revisão Bibliográfica

As Redes Neurais Convolucionais foram inicialmente propostas por LeCun et al. na década de 1990, com a introdução do modelo LeNet para reconhecimento de dígitos manuscritos. Desde então, avanços significativos foram alcançados, especialmente com a introdução de arquiteturas mais profundas, como AlexNet e VGGNet, que mostraram um desempenho sem precedentes em competições como o ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

- **Szeliski (2010):** Discutiu algoritmos clássicos de visão computacional, fornecendo uma base para o entendimento de CNNs.
- **He et al. (2016):** Apresentaram a arquitetura ResNet, que introduziu blocos residuais, resolvendo o problema de degradação em redes profundas.
- **Goodfellow et al. (2016):** Descreveram o aprendizado profundo de maneira abrangente, incluindo a teoria e aplicações das CNNs.
- **Krizhevsky et al. (2012):** Demonstraram a eficácia das CNNs com o AlexNet, que reduziu significativamente os erros no desafio ImageNet.

Além disso, arquiteturas mais recentes, como o EfficientNet e o Vision Transformer, têm demonstrado avanços significativos em tarefas de visão computacional. Essas arquiteturas utilizam abordagens inovadoras, como o escalonamento eficiente de profundidade, largura e resolução no caso do EfficientNet, e o uso de atenção no caso do Vision Transformer, permitindo maior precisão e eficiência em datasets de grande escala.

III. Materiais e Métodos

Nesta seção de Materiais e Métodos, são descritos detalhadamente os procedimentos e ferramentas utilizados para a execução deste estudo. Isso inclui a seleção e preparação do conjunto de dados, as arquiteturas de redes neurais empregadas, e as configurações experimentais para o treinamento e avaliação dos modelos. O objetivo principal é oferecer uma visão clara e replicável das etapas que fundamentaram o experimento, garantindo a integridade dos resultados obtidos e sua comparação com futuros trabalhos na área. A seguir, são apresentados os principais elementos utilizados neste estudo.

A. Dataset

O conjunto escolhido foi o Animals10, um dataset que contém imagens de 10 classes de animais, como gatos, cães, cavalos, entre outros. As imagens foram redimensionadas para 224×224 pixels e normalizadas com os valores médios e desvios padrão do conjunto ImageNet. A divisão do dataset foi realizada da seguinte forma:

- **Conjunto de Treinamento:** 80% das imagens.
- **Conjunto de Validação:** 20% das imagens restantes.

Essa divisão foi estratificada para garantir uma distribuição proporcional entre as classes.

B. Modelos Utilizados

Para este estudo, foram selecionados dois modelos predefinidos do PyTorch, conhecidos por suas características e desempenho em tarefas de classificação de imagens:

- **AlexNet:** Desenvolvido para o desafio ImageNet, o AlexNet foi um dos primeiros modelos a demonstrar o potencial das redes neurais convolucionais em grande escala. Ele utiliza uma arquitetura relativamente simples, com camadas convolucionais e totalmente conectadas. Neste trabalho, a camada de saída do AlexNet foi ajustada para conter 10 neurônios, correspondendo ao número de classes do dataset Animals10.
- **ResNet-50:** Uma rede mais moderna e profunda, a ResNet-50 utiliza blocos residuais para facilitar o treinamento de redes com muitas camadas. Os blocos residuais ajudam a mitigar problemas de degradação de desempenho, comuns em redes profundas. No contexto deste estudo, a camada final foi modificada para lidar com as 10 classes do Animals10.

Além disso, foi desenvolvida uma CNN personalizada para comparação de desempenho. A arquitetura dessa rede inclui:

- Duas camadas convolucionais com função de ativação ReLU e operações de pooling.
- Três camadas totalmente conectadas, projetadas para a classificação final.

Os modelos AlexNet e ResNet-50 foram utilizados com pesos pré-treinados no ImageNet, o que permitiu inicializar os parâmetros com representações úteis, reduzindo o tempo de treinamento necessário para alcançar alta precisão. A camada de saída de ambos os modelos foi ajustada para corresponder às 10 classes do dataset Animals10.

C. Treinamento e Avaliação

O treinamento e a avaliação dos modelos seguiram uma abordagem meticulosa, utilizando o framework PyTorch. As etapas principais incluíram a configuração de hiperparâmetros, o uso de validação cruzada e a análise detalhada das métricas de desempenho.

Hiperparâmetros do Treinamento:

- **Taxa de aprendizado:** 0.001.
- **Otimizador:** SGD com momento de 0.9.
- **Função de perda:** CrossEntropyLoss, escolhida por sua eficácia em problemas de classificação multiclasses.
- **Épocas:** 50.
- **Tamanho do lote:** 64, balanceando eficiência computacional e convergência estável.

O processo de treinamento envolveu a atualização iterativa dos pesos do modelo usando o método de retropropagação, com os dados processados em lotes para otimizar o uso de memória e desempenho. A função de perda CrossEntropyLoss foi utilizada para calcular o erro entre as predições do modelo e os rótulos verdadeiros.

Os experimentos foram realizados em uma GPU T4, com tempos médios de treinamento de aproximadamente 97 minutos para o AlexNet e 272 minutos para o ResNet-50. Esse ambiente computacional utilizado através do Googe Colab foi essencial para acelerar o processo de treinamento e validar o desempenho das arquiteturas.

Estratégia de Avaliação:

- Após cada época de treinamento, os modelos foram avaliados no conjunto de validação para monitorar a perda e a acurácia.
- Para cada batch, as predições foram comparadas aos rótulos verdadeiros, permitindo o cálculo de métricas como:
  - **Acurácia:** Proporção de predições corretas em relação ao total.
  - **Precisão:** Percentual de predições positivas corretas.
  - **Recall:** Capacidade do modelo de identificar todas as instâncias positivas.
  - **F1-Score:** Média harmônica entre precisão e recall.
- Matrizes de confusão foram geradas para entender os padrões de erro e identificar possíveis melhorias no modelo.

Pipeline do Treinamento: O pipeline incluiu as seguintes etapas:

- Habilitação do modo de treinamento do modelo (`model.train()`).
- Forward pass: Computação das predições do modelo.
- Backward pass: Cálculo do gradiente e atualização dos pesos.
- Avaliação no conjunto de validação utilizando `model.eval()`, desabilitando o cálculo de gradientes para melhorar a eficiência.

Gráficos de desempenho foram gerados ao final do treinamento para visualizar a evolução das perdas e acurácias durante as épocas. Essas informações são cruciais para entender o comportamento do modelo e identificar possíveis pontos de ajuste nos hiperparâmetros ou na arquitetura.

IV. Resultados e Discussão

A. Desempenho dos Modelos

Os resultados quantitativos dos modelos AlexNet e ResNet-50 no conjunto de validação estão resumidos na Tabela I. O ResNet-50 obteve desempenho superior em todas as métricas avaliadas, com uma acurácia geral de 98.36%, enquanto o AlexNet alcançou uma acurácia de 93.98%. Isso reflete a maior capacidade do ResNet-50 em capturar características

mais complexas devido à sua arquitetura profunda e uso de blocos residuais.

B. Matrizes de Confusão

As matrizes de confusão dos modelos, apresentadas nas Figuras 1 e 2, revelam a capacidade de cada modelo em classificar corretamente as 10 classes de animais no conjunto de validação. O AlexNet mostrou dificuldades em classes mais complexas, como *mucca* e *pecora*, enquanto o ResNet-50 apresentou maior precisão em todas as classes, evidenciando sua robustez.

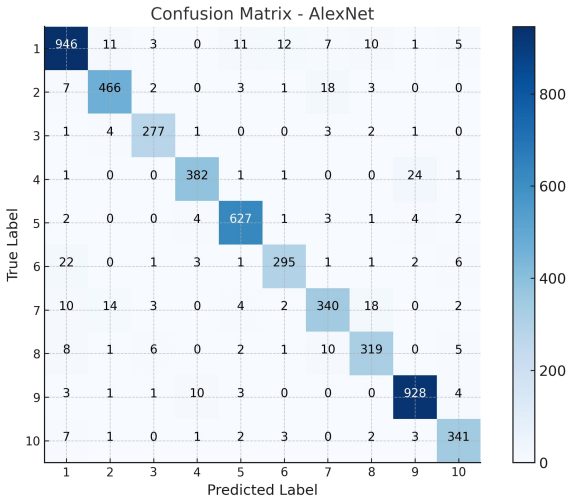

Fig. 1. Matriz de confusão do AlexNet no conjunto de validação.

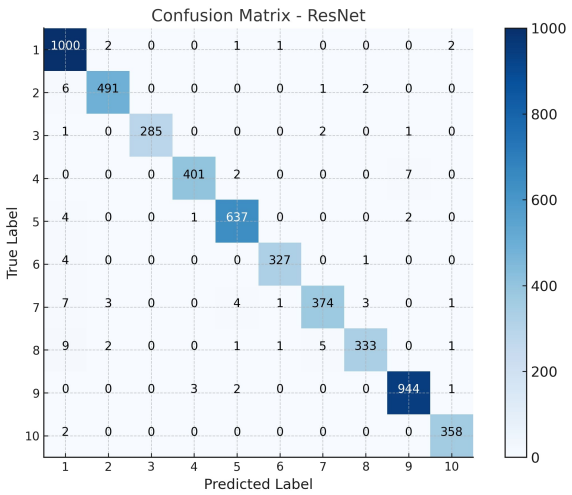

Fig. 2. Matriz de confusão do ResNet-50 no conjunto de validação.

C. Curvas de Aprendizado

As curvas de perda e acurácia para os modelos AlexNet e ResNet-50 estão representadas nas Figuras 3, 4 e 5, 6, respectivamente. As curvas do ResNet-50 indicam uma convergência mais rápida e consistente, enquanto o AlexNet apresenta flutuações na perda de validação, sugerindo que o modelo pode estar mais suscetível a overfitting.

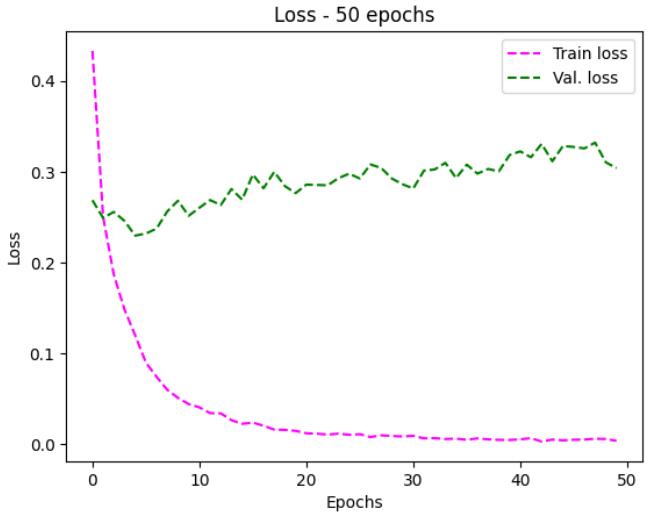

Fig. 3. Curvas de perda do AlexNet.

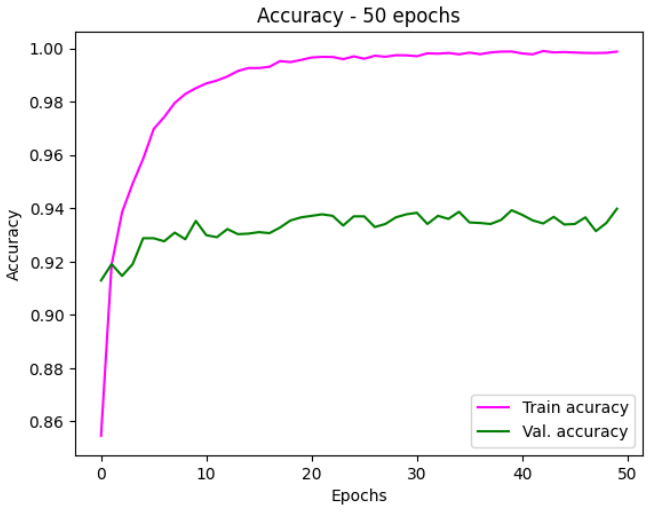

Fig. 4. Curvas de acurácia do AlexNet.

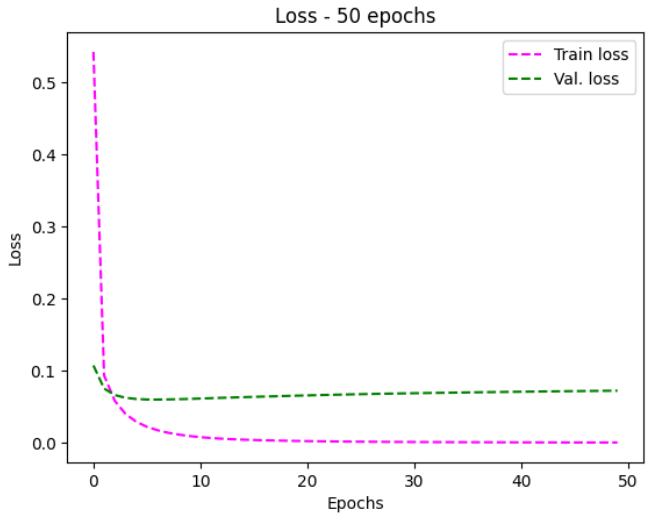

Fig. 5. Curvas de perda do ResNet-50.

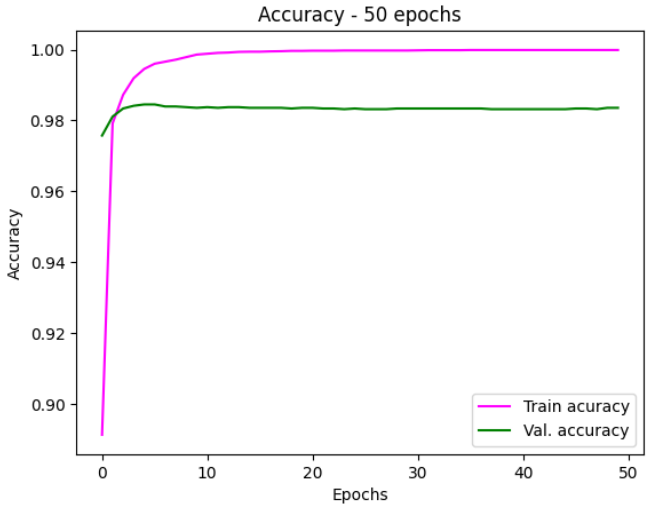

Fig. 6. Curvas de acurácia do ResNet-50.

D. Discussão dos Resultados

Os resultados destacam o impacto das arquiteturas mais modernas, como a ResNet-50, na melhoria do desempenho em tarefas de classificação de imagens. O uso de blocos residuais no ResNet-50 permitiu uma melhor generalização, enquanto o AlexNet, apesar de seu desempenho respeitável, foi limitado por sua simplicidade arquitetural. As análises realizadas demonstram a importância de selecionar arquiteturas apropriadas para lidar com a complexidade do dataset.

Além disso, as matrizes de confusão evidenciam que, embora o ResNet-50 seja superior, ambos os modelos apresentaram erros concentrados em classes com características visuais semelhantes, como *mucca* e *pecora*. Esses resultados sugerem que técnicas adicionais, como aumento de dados ou treinamento com arquiteturas mais avançadas, podem melhorar ainda mais o desempenho.

E. Análise das Métricas

As métricas de avaliação utilizadas incluem:

- **Precisão (Precision):** Mede a proporção de predições positivas corretas em relação ao total de predições positivas.
- **Recall:** Mede a capacidade do modelo de identificar corretamente todas as instâncias positivas.
- **F1-Score:** Representa a média harmônica entre precisão e recall, sendo especialmente útil quando há um desequilíbrio nas classes.
- **Acurácia:** Reflete a proporção de predições corretas em relação ao total de amostras.

Essas métricas fornecem uma visão abrangente do desempenho do modelo, avaliando sua capacidade de generalizar e lidar com classes desafiadoras.

A análise das métricas de desempenho, como *F1-score*, *recall*, *precisão* e *acurácia*, forneceu insights importantes sobre a eficácia dos modelos AlexNet e ResNet-50 na tarefa de classificação de imagens. Essas métricas foram avaliadas no conjunto de validação, com foco nas 10 classes do dataset Animals10.

O modelo AlexNet apresentou um *F1-score* médio ponderado de 93.97%, com precisão (*precision*) e *recall* também de 93.97%. Apesar de um bom desempenho geral, algumas classes, como *mucca* e *pecora*, mostraram menores valores individuais de *F1-score* e *recall*, indicando dificuldade do modelo em diferenciar essas categorias. Isso pode ser atribuído à semelhança visual entre os animais dessas classes, além de possíveis limitações na arquitetura do AlexNet, que é relativamente simples em comparação com arquiteturas mais modernas.

Por outro lado, o ResNet-50 apresentou métricas significativamente superiores, com um *F1-score* médio ponderado de 98.35%, precisão (*precision*) de 98.37% e *recall* de 98.36%. Esses valores indicam que o ResNet-50 é um modelo mais robusto, capaz de capturar características mais complexas graças ao uso de blocos residuais, que mitigam o problema de degradação de gradientes em redes profundas. O desempenho do ResNet-50 foi consistentemente alto em todas as classes, destacando-se especialmente nas classes *cavallo*, *ragno* e *scoiattolo*, com *F1-scores* acima de 98.5%.

A acurácia geral dos modelos reforça essa análise: o AlexNet obteve uma acurácia de 93.98%, enquanto o ResNet-50 alcançou uma acurácia de 98.36%, consolidando sua superioridade. As diferenças nas métricas individuais e gerais refletem a capacidade do ResNet-50 de generalizar melhor em classes visualmente desafiadoras, enquanto o AlexNet, embora competente, é limitado pela simplicidade de sua arquitetura.

Esses resultados destacam a importância de selecionar arquiteturas modernas e avançadas para tarefas complexas de classificação de imagens, como as encontradas no dataset Animals10. Além disso, técnicas adicionais, como aumento de dados ou regularização, poderiam ser exploradas para melhorar ainda mais o desempenho do AlexNet e aproximá-lo do ResNet-50 em termos de robustez.

A Tabela I resume os valores de *F1-score*, *recall* e *precisão* para ambos os modelos. Como esperado, o ResNet-50 superou o AlexNet em todas as métricas avaliadas, demonstrando uma maior capacidade de generalização e precisão na classificação.

TABLE I
MÉTRICAS DE DESEMPENHO DOS MODELOS NO CONJUNTO DE VALIDAÇÃO.

Modelo	Acurácia	Precisão (w)	Recall (w)	F1-Score (w)
AlexNet	93.98%	93.97%	93.98%	93.97%
ResNet-50	98.36%	98.37%	98.36%	98.35%

Outra métrica importante é a *acurácia*, que reflete a proporção de predições corretas em relação ao total de amostras. O ResNet-50 alcançou uma acurácia de 98.36%, enquanto o AlexNet obteve 93.98%. Essa diferença reflete a capacidade do ResNet-50 de lidar melhor com classes minoritárias e com maior variabilidade de padrões nas imagens.

No caso do AlexNet, a matriz de confusão revelou maior taxa de confusão entre classes como *mucca* e *pecora*, além de erros ocasionais em *farfalla* e *gatto*. Em contraste, o ResNet-50 apresentou erros mínimos, com alta precisão em quase todas as classes, demonstrando maior estabilidade e generalização.

Por fim, a consistência do *recall* e do *F1-score* para o ResNet-50 reforça sua adequação para aplicações em que a identificação de todas as instâncias relevantes é crucial. Embora o AlexNet tenha mostrado um desempenho respeitável, suas limitações estruturais tornam-no mais suscetível a erros em classes com características visuais semelhantes. O uso de arquiteturas mais modernas, como o ResNet-50, é, portanto, recomendado para tarefas mais exigentes de classificação de imagens.

V. CONCLUSÃO

Os resultados deste estudo demonstram de forma clara a superioridade de arquiteturas modernas, como a ResNet-50, na tarefa de classificação de imagens em datasets desafiadores como o Animals10. O desempenho do ResNet-50, com uma acurácia de validação de 98.36%, superou amplamente o AlexNet, que alcançou 93.98%. Esses resultados são consistentes com as vantagens esperadas do ResNet-50, incluindo o uso de blocos residuais que permitem um treinamento mais eficiente e uma melhor generalização, especialmente em classes com características visuais semelhantes.

O AlexNet, por sua vez, apresentou um desempenho respeitável, com métricas competitivas e uma acurácia consistente ao longo das épocas. No entanto, suas limitações estruturais resultaram em maior confusão entre classes com padrões similares, como *mucca* e *pecora*, destacando a necessidade de arquiteturas mais robustas para tarefas mais complexas.

As matrizes de confusão e as curvas de aprendizado reforçaram os resultados quantitativos, oferecendo uma visão mais profunda sobre os erros cometidos por cada modelo. Enquanto o AlexNet mostrou flutuações na perda de validação, sugerindo uma leve tendência ao *overfitting*, o ResNet-50 apresentou curvas mais consistentes e um aprendizado rápido, convergindo para valores estáveis após poucas épocas.

Os resultados também evidenciam a importância da seleção adequada de arquiteturas para tarefas específicas. Em aplicações em que a precisão e a generalização são cruciais, como diagnóstico médico automatizado ou sistemas de segurançã, modelos mais complexos, como o ResNet-50, são claramente a melhor escolha. No entanto, em cenários onde a simplicidade e a eficiência computacional são prioritárias, o AlexNet ainda pode oferecer uma solução viável.

Este estudo também reforça a relevância das Redes Neurais Convolucionais em aplicações reais, destacando o impacto de avanços arquiteturais no desempenho de modelos de classificação de imagens. Além disso, as análises das métricas e das matrizes de confusão fornecem uma base sólida para pesquisadores interessados em otimizar ainda mais essas arquiteturas.

Como trabalhos futuros, sugere-se a exploração de técnicas complementares, como aumento de dados (*data augmentation*) para aumentar a diversidade do dataset, e *fine-tuning* de hiperparâmetros para refinar ainda mais o desempenho dos modelos. Além disso, a implementação de redes ainda mais profundas, como o EfficientNet ou o Vision Transformer, poderia trazer insights adicionais sobre a evolução das arquiteturas de redes neurais em tarefas de visão computacional.

Os resultados reforçam a aplicabilidade de arquiteturas modernas como o ResNet-50 em problemas complexos, oferecendo soluções robustas para aplicações críticas, como diagnósticos médicos e sistemas de segurança. Além disso, este trabalho destaca a importância do uso de recursos computacionais adequados para explorar todo o potencial de arquiteturas profundas, como o ResNet-50.

Para estudos futuros, recomenda-se a implementação de técnicas avançadas, como aumento de dados (*data augmentation*) para aumentar a diversidade do dataset e *fine-tuning* para ajustar os hiperparâmetros. Arquiteturas ainda mais modernas, como o Vision Transformer e o EfficientNet, podem ser exploradas para expandir os insights apresentados neste estudo.

Por fim, este trabalho contribui para o entendimento e aplicação de Redes Neurais Convolucionais, evidenciando sua importância em soluções baseadas em aprendizado profundo e destacando as melhores práticas para futuros estudos no campo.

REFERENCIAS

[1] K. He et al., "Deep Residual Learning for Image Recognition," *CVPR*, 2016.
[2] I. Goodfellow et al., *Deep Learning*, MIT Press, 2016.
[3] A. Krizhevsky et al., "ImageNet Classification with Deep Convolutional Neural Networks," *NIPS*, 2012.
[4] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2010.