

# Detecção de Pneumonia (Classificação de Imagens Médicas)

Arthur Gabriel Silva da Luz, UFPE ([agsl@cin.ufpe.br](mailto:agsl@cin.ufpe.br))

**Resumo**—A pneumonia é uma das principais causas de mortalidade por doenças infecciosas no mundo, tornando essencial o desenvolvimento de métodos automatizados que auxiliem no diagnóstico precoce. Este trabalho apresenta uma solução baseada em Redes Neurais Convolucionais (CNNs) para a classificação binária de radiografias de tórax em duas categorias: Normal e Pneumonia. Foi utilizada uma base de dados composta por 5232 imagens para treinamento e validação, com aplicação de técnicas de data augmentation e transfer learning. Esse relatório apresenta duas implementações de redes neurais convolucionais (RNCs): ResNet50 e ViT-16. A Resnet-50 apresentou melhor desempenho, alcançando valores elevados de precisão, recall e ROC-AUC com alta capacidade de separação entre as classes. A técnica de interpretabilidade Grad-CAM confirmou que as ativações se concentraram em áreas pulmonares compatíveis com padrões patológicos e os resultados indicam que a solução proposta possui elevado potencial como ferramenta de apoio à decisão clínica.

## I. INTRODUÇÃO

Segundo a Organização Mundial de Saúde(OMS), a pneumonia é uma doença que ceifa a vida de mais de 700.000 crianças por ano e é considerada a maior causa de morte dentre as doenças infecciosas.

O diagnóstico precoce e preciso é fundamental para um tratamento eficaz e para a redução das taxas de mortalidade. Atualmente, este diagnóstico baseia-se na avaliação clínica e na análise de radiografias de tórax(raio-x) por radiologistas especializados. No entanto, a interpretação dessas imagens está suscetível a erros humanos e limitada pela disponibilidade de especialistas, especialmente em regiões com recursos escassos.

Hoje em dia, a tecnologia avançou a ponto de poder ser utilizada como auxílio em diversas áreas, dentre elas, a saúde. A tarefa de identificar padrões em imagens, o que é natural para o ser humano, conseguiu ser replicada com sucesso por modelos computacionais via redes neurais de múltiplas camadas. Essa vertente da aprendizagem de máquina é conhecida como deep learning e baseia-se em modelos matemáticos inspirados no cérebro humano, nos quais neurônios artificiais processam dados e propagam seus resultados entre esses neurônios [1].

A eficácia dos modelos convolucionais, potencializadas a partir de 2012 com o pioneirismo da rede neural convolucional(CNN) AlexNet[2] permitiu que o reconhecimento de imagens evoluísse muito e alcançasse alta precisão em diversas áreas do conhecimento, dentre elas, os diagnósticos médicos.

Deste modo, este relatório apresenta o desenvolvimento de uma solução baseada em Redes Neurais Convolucionais (CNNs) para a detecção de pneumonia ou não em imagens de raio-x. O propósito é diminuir a sobrecarga nos

profissionais médicos e garantir um diagnóstico coerente através de uma automação de uma classificação binária entre pulmões saudáveis(Normal) e patológicos(Pneumonia).

## II. BASE DE DADOS

A base de dados utilizada foi fornecida pela Liga de Inteligência Artificial(Ligia) e disponibilizada na plataforma Kaggle. Esse conjunto de dados é composto por 5232 imagens de raio-x de tórax para treino, organizado em duas classes(1349 imagens Normal e 3883 Pneumonia) mais 624 imagens de teste sem rótulo.

### A. Data Augmentation e pré-processamento de dados

Data Augmentation é uma estratégia de regularização em CNNs que possui o objetivo de mitigar o overfitting por meio do aumento sintético do conjunto de treinamento. Por meio dessa técnica, o modelo é treinado com diversos exemplos sintéticos, consequentemente obtendo um melhor aprendizado.

Dessa forma, o aumento de dados impede que a arquitetura memorize ruídos específicos presentes no dataset original. Logo, há uma redução significativa no erro de generalização do modelo e há menos chances dele se tornar um especialista apenas nas imagens do conjunto de treino.

As imagens foram submetidas a um pipeline de pré-processamento por meio da biblioteca TorchVision. Todas as amostras foram redimensionadas para a resolução espacial de 224x224 pixels e normalizadas utilizando as médias e desvio-padrão do conjunto ImageNet (médias: [0.485, 0.456, 0.406]; desvios: [0.229, 0.224, 0.225]) e mais algumas configurações:

- Batch\_size = 32
- Otimizador = Adam
- Epochs = 30
- Learning\_rate = 0.0001
- transforms.Resize((224, 224)),
- transforms.RandomResizedCrop(224,scale=(0.8, 1.0)),
- transforms.RandomHorizontalFlip(),
- transforms.RandomRotation(10),
- transforms.ColorJitter(brightness=0.2,contrast=0.2)
- transforms.ToTensor(),
- transforms.Normalize(mean=médias, std=desvios)

### B. Robustez e reprodutibilidade



Para garantir a robustez e reprodutibilidade da avaliação, os dados foram particionados aleatoriamente em três subconjuntos diferentes, com uma semente fixa(seed=42) para reprodutibilidade:

Treinamento (70%): Divisão para ajuste dos pesos do modelo

Validação (30%): Utilizado para o monitoramento durante o treinamento e ajuste dos hiperparâmetros.

### III. METODOLOGIA

#### A. Redes Neurais Convolucionais

As redes neurais convolucionais(CNNs) representam um avanço na área de inteligência artificial, utilizando princípios que foram inspirados no cérebro humano.

Diferente dos modelos estatísticos genéricos, essas estruturas utilizam deep learning para interpretar dados com grande volume e multidimensionais. A principal vantagem reside na capacidade de reconhecer padrões, mesmo diante de alterações no contraste, brilho, rotação ou escala de imagem.

As técnicas de visão computacional clássicas dependiam de filtros manuais, já as CNN's automatizam a detecção de características(features extraction) através de camadas de convolução[3]. Nesta etapa, a rede consegue identificar as correlações espaciais entre pixels adjacentes através de “máscaras matemáticas”(kernels) que “deslizam” sobre a imagem da esquerda para a direita e de cima para baixo.

Após a feature extraction primária, as camadas de pooling vão destilar a informação. A partir disso, realiza-se uma subamostragem e a rede consegue diminuir a resolução espacial dos dados, o que torna o modelo mais eficiente computacionalmente, mas também confere à rede uma maior tolerância a pequenas distorções geométricas nos dados de entrada.

Na fase final, o conhecimento extraído é enviado para as camadas totalmente conectadas, que funcionam como um classificador final e conseguem integrar todos os sinais recebidos para realizar a inferência.

O aprimoramento desses modelos CNN's ocorre por meio de um algoritmo chamado backpropagation(retropropagação). Esse algoritmo calcula os erros e a otimização via gradiente descendente e ajusta os pesos da rede, permitindo que a rede neural consiga distinguir padrões complexos nos exames médicos.

Por fim, para atender ao requisito estabelecido pela organização, as saídas do modelo contidas no intervalo [0,1] foram convertidas por meio da função Softmax para transformar os valores brutos em uma distribuição de probabilidade e por conseguinte facilitando a interpretação de cada valor predito.

#### B. Modelos de Redes neurais Convolucionais

- **Resnet50:** Uma variante das Redes Neurais convolucionais(CNN's) que introduziu o conceito de aprendizado residual para viabilizar o treino de redes profundas. Essa arquitetura se diferencia por meio da implementação de skip connections ou shortcut connections(Conexões residuais) que permitem que o sinal pule camadas intermediárias durante o processamento.

- **Vision Transformers(VIT\_16):** Representam uma evolução no âmbito da visão computacional ao aplicar a arquitetura dos transformers diretamente a dados visuais, não dependendo mais exclusivamente de operações convolucionais. A imagem é dividida em pequenos blocos, linearmente projetado e a arquitetura utiliza o mecanismo de self-attention para que o modelo aprenda as partes importantes entre as diferentes imagens.

#### C. Estratégia de modelagem e Transfer-Learning

##### 1) Fine-tuning parcial e Congelamento de camadas

Dada a especificidade das imagens médicas, adotou-se uma estratégia de fine-tuning parcial, na qual o código percorre os parâmetros do modelo e aplica um congelamento seletivo:

- **Camadas Iniciais e Intermediárias:** Foram mantidas congeladas preservando as features extractions de baixo nível

- **Camadas de Alto nível e Classificador:** O bloco residual (Layer4) e a camada totalmente conectada(fc) foram mantidos treináveis. Essa decisão se baseou no fato de que camadas mais profundas extraem representações mais complexas, que precisam ser adaptadas de acordo com os padrões nas radiografias de pneumonia.

##### 2) Reconfiguração da Camada de Classificação

A camada de classificação da Resnet-50 foi substituída por uma estrutura sequencial personalizada composta por:

1. **Camada Linear Intermediária:** Reduz a dimensionalidade para 512 neurônios.

2. **Batch Normalization (BatchNorm1d):** Aplicado para estabilizar a distribuição das ativações, acelerar a convergência e fornecer um leve efeito de regularização.

3. **Função de Ativação ReLU:** Introduce a não-linearidade necessária para o aprendizado de padrões complexos.

4. **Dropout (0.3):** Camada de regularização que desativa aleatoriamente 30% dos neurônios durante o treinamento, usado para prevenir o *overfitting* e aumentar a capacidade de generalização.

5. **Camada de Saída:** Uma camada linear final com 2 neurônios, correspondentes às classes "Normal" e "Pneumonia".

## IV. RESULTADOS

### Matriz de confusão ResNet-50

#### A. Métricas de avaliação

1) Precisão: Avalia a exatidão das classificações positivas realizadas por um modelo. Matematicamente é definida como a razão entre os verdadeiros positivos(VP) e a soma entre falsos positivos(FP) e verdadeiros positivos através da fórmula:

$$\text{Precisão} = VP / (VP + FP)$$

2) Recall: Mensura a capacidade do modelo criado em rastrear as instâncias relevantes presentes no conjunto de dados. Um alto recall é vital pois garante que a rede neural treinada está deixando passar a menor quantidade possível de imagens positivas, o que minimiza os falsos negativos(FN).

$$\text{Recall} = VP / (VP + FN)$$

3) Roc-auc: A curva ROC mapeia o balanço entre a taxa de acertos(Verdadeiros Positivos) e a taxa de alarmes falsos(Falsos Positivos) à medida que esse limiar da decisão varia. Quanto mais próximo de 1,0 tiver o valor de AUC, maior a robustez do modelo.

TABELA 1.  
Performance dos modelos

Model	Classe	Precisão	ROC-AUC	Recall
Resnet50	Normal	0.95	0.9979	0.97
	Pneumonia	0.99	0.9979	0.98
Vision Transformers (Vit_16)	Normal	0.91	0.9945	0.96
	Pneumonia	0.99	0.9945	0.96

Com base nos resultados apresentados na Tabela 1, a arquitetura ResNet-50 foi selecionada como a abordagem principal para o problema de classificação. Por conseguinte, as avaliações de desempenho e as visualizações gráficas subsequentes foram desenvolvidas a partir deste modelo.

A análise dos erros cometidos revela informações acerca da aplicabilidade clínica da solução. A ocorrência de 12 falsos negativos, indica 12 casos em que o modelo falhou em detectar a patologia presente. Este é o erro mais crítico que pode-se ter, pois indica a ausência de tratamento para um paciente infectado.

O outro ponto a ser analisado é que o modelo gerou 18 falsos positivos, ou seja 18 pacientes saudáveis seriam classificados como doentes. Esses casos apresentam um risco clínico menor em comparação a falsos negativos, mas ainda exigem dos pacientes custos adicionais com exames confirmatórios e até medicamentos não necessários.

Por fim, a predominância de acertos indica que a estratégia de pesos na loss function (função de perda) foi eficaz para obter uma alta detecção da condição patológica(pneumonia).

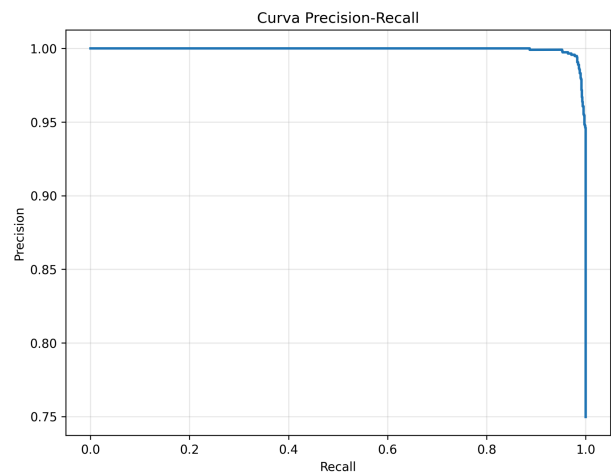


FIGURA 2:  
Precision-Recall (PR Curve)

A figura 2 apresenta uma manutenção de altos níveis de precisão mesmo à medida que o recall aumenta, isso demonstra que o modelo é capaz de identificar uma grande parte dos casos de pneumonia (minimiza falsos negativos), sem ter um aumento de diagnósticos incorretos para pacientes saudáveis (falsos positivos).

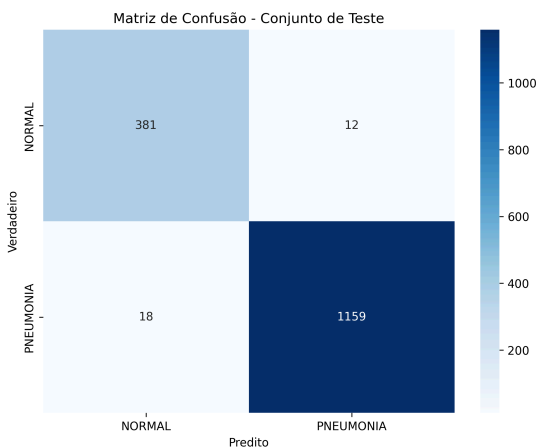


FIGURA 1:

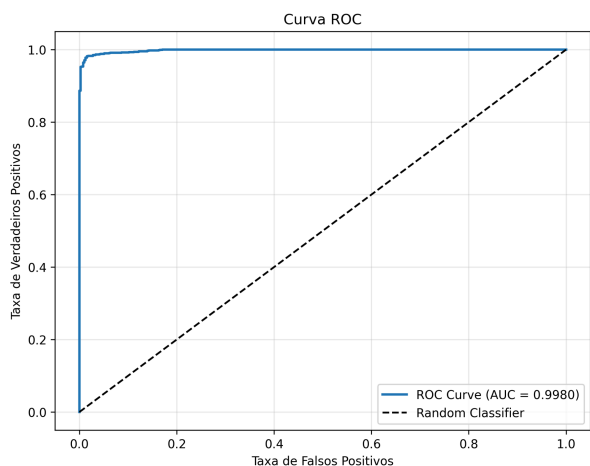


FIGURA 3:  
Curva ROC-AUC

A figura 3 apresenta a Curva de Característica da Operação do Receptor (ROC) avaliada no conjunto de teste. A curva aproxima-se idealmente do canto superior esquerdo, culminando em uma área sob a curva (AUC) de 0.9980. Este valor, muito superior à linha de base de predição aleatória (0.5) evidencia que a rede neural ResNet-50 teve sucesso em separar as distribuições das classes “Normal” e “Pneumonia”.

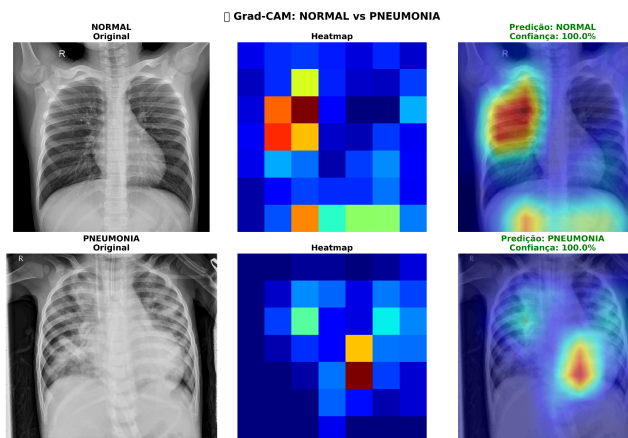


FIGURA 4:  
Mapa de Calor Grad-CAM

A figura 4 apresenta a visualização por meio do algoritmo de inteligência artificial explicável Grad-Cam(Gradient-Weighted Class Activation Mapping), um método de interpretabilidade para RNC's que permite identificar quais regiões da imagem o modelo prestou mais atenção para tomar a decisão final(A cor vermelha indica uma região de maior atenção, amarelo uma região moderadamente importante, verde indica uma região pouco importante e o azul é a região ignorada pelo modelo).

## V. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Apesar dos resultados expressivos, algumas limitações devem ser consideradas para análise da solução. Primeiramente, há um desbalanceamento no conjunto de dados (3883 imagens de pneumonia contra 1349 imagens normais), o que pode introduzir um viés no treinamento,

mesmo com as estratégias utilizadas de ponderação na função de perda.

Outra limitação refere-se ao fato de que o modelo foi treinado e validado dentro de um mesmo domínio de dados, o que pode limitar sua robustez perante diferentes equipamentos e protocolos.

Os resultados indicam que a abordagem proposta apresenta um grande potencial como **ferramenta de apoio** ao diagnóstico com capacidade de auxiliar profissionais de saúde, reduzindo a sobrecarga clínica.

Por fim, como trabalho futuro, algumas melhorias que poderiam ser feitas incluem a expansão do conjunto de dados, exploração de diferentes arquiteturas, implementação de outras classes patológicas e aplicação de outras técnicas avançadas de explicabilidade ou métodos baseados em atenção(Transformers).

## VI. REFERÊNCIAS

- [1]: SILVA, Rodrigo Emerson Valentim da. Um estudo comparativo entre redes neurais convolucionais para a classificação de imagens. 2018. 51 f. TCC (Graduação em Sistemas de Informação) Universidade Federal do Ceará, Campus de Quixadá, Quixadá, 2018.
- [2]: Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNetClassification with Deep Convolutional Neural Networks. In Advances inNeural Information Processing Systems 25: 26th Annual Conference on NeurallInformation Processing Systems 2012. Proceedings of a meeting held Decem-ber 3-6, 2012, Lake Tahoe, Nevada, United States, Peter L. Bartlett, FernandoC. N. Pereira, Christopher J. C. Burges, Léon Bottou, and Kilian Q. Wein-berger (Eds.). 1106–1114.
- [3]: Wafa Mousser and Salima Ouadfel. 2019. Deep Feature Extraction for Pap-Smear Image Classification: A Comparative Study. In Proceedings of the 2019 5th International Conference on Computer and Technology Applications (ICCTA 2019). ACM, New York, NY, USA, 6-10. DOI: <https://doi.org/10.1145/3323933.3324060>

