

Estudo Comparativo de Algoritmos de Classificação de Imagens na Identificação de Pneumonia em Raios-x Pulmonares

Jennifer Mayara de Paiva Goberski
Escola de Tecnologia da Informação
Universidade Positivo
Curitiba, Brasil
jenni.goberski@gmail.com

João Vitor Maciel de Brito
Escola de Tecnologia da Informação
Universidade Positivo
Curitiba, Brasil
joao.vmacielb@gmail.com

Marcel Antunes Raposo
Escola de Tecnologia da Informação
Universidade Positivo
Curitiba, Brasil
marcel.raposo@cs.up.edu.br

Vinicius Dionizio Patrocínio
Escola de Tecnologia da Informação
Universidade Positivo
Curitiba, Brasil
vinipatrocínio.vd@gmail.com

Leonardo Gomes Tavares
Escola Politécnica
Universidade Positivo
Curitiba, Brasil
leonardo.tavares@up.edu.br

Resumo—Este trabalho apresenta uma análise comparativa de diferentes modelos de técnicas de aprendizado de máquina para a classificação binária de imagens de raio-x pulmonares a fim de identificar a presença ou ausência de pneumonia. Utilizando uma base de dados pré-processada, que incluiu normalização e balanceamento de classes, foram aplicadas técnicas de *data augmentation* e pesos de classe para otimizar o treinamento dos modelos. Entre os modelos clássicos avaliados, como Regressão Logística, Árvore de Decisão, SVM, MLP e *Random Forest*, a SVM destacou-se com a melhor performance, obtendo uma pontuação média de *F1 score* de 0,96. No entanto, os modelos de aprendizado profundo, particularmente as Redes Neurais Convolucionais (CNNs), como *ResNet* e *EfficientNet*, superaram significativamente os modelos tradicionais, alcançando *F1 scores* de até 0,98. A *ResNet* foi identificada como a arquitetura mais eficaz, graças à sua capacidade de capturar características complexas nas imagens de raio-x. Os resultados sugerem que o uso desses modelos pode aprimorar a precisão diagnóstica, oferecendo suporte valioso para profissionais de saúde na detecção precoce de doenças pulmonares.

Abstract—This work presents a comparative analysis of different machine learning models for binary classification of chest X-ray images to identify the presence or absence of pneumonia. Utilizing a pre-processed dataset, which included normalization and class balancing, data augmentation techniques and class weights were applied to optimize the training of the models. Among the classic models evaluated, such as Logistic Regression, Decision Tree, SVM, MLP and Random Forest, SVM demonstrated the best performance, achieving an average F1 score of 0.96. However, deep learning models, particularly Convolutional Neural Networks (CNNs) like ResNet and EfficientNet, significantly outperformed traditional models, reaching F1 scores of up to 0.98. ResNet was identified as the most effective architecture, owing to its ability to capture complex features in X-ray images. The results suggest that the use of these models can enhance diagnostic accuracy, providing valuable support to healthcare professionals in the early detection of lung diseases.

Palavras-chave—Aprendizado de máquina, aprendizado profundo, pneumonia, raio-x, diagnóstico.

I. INTRODUÇÃO

A pneumonia adquirida na comunidade (PAC) é uma das principais causas de morte em todo o mundo, especialmente entre crianças menores de cinco anos [4], [15]. Vários fatores, como baixo peso ao nascer, nascimento prematuro, desnutrição e vacinação incompleta, podem estar associados a essa taxa de mortalidade infantil [3].

Clinicamente, a PAC manifesta-se com sintomas como febre, tosse e sinais de infecção no trato respiratório inferior, sendo caracterizada como uma infecção adquirida fora do ambiente hospitalar [3]. Para o diagnóstico, podem ser utilizados métodos como tomografia computadorizada, ressonância magnética e radiografias (raio-x), sendo este último o método mais acessível e amplamente disponível para obtenção de imagens pulmonares [11].

Devido à complexidade da patogenicidade da PAC, à resistência aos antibióticos e à incerteza diagnóstica, seu manejo clínico apresenta desafios significativos [15]. Esses fatores destacam a importância de pesquisas voltadas para a melhoria dos métodos de detecção, especialmente em estágios iniciais da doença, para evitar que ela progrida para um estado potencialmente fatal [10].

Com o avanço das técnicas de processamento de imagens, tanto tradicionais quanto as baseadas em aprendizado de máquina, pesquisadores têm explorado essas tecnologias na medicina para a detecção de diversas doenças, incluindo câncer de pele, câncer de mama, tumores cerebrais, tuberculose e pneumonia [11].

Neste contexto, o presente trabalho realiza uma análise comparativa entre diferentes modelos de aprendizado de máquina tradicionais (*traditional machine learning*) e aprendizado profundo (*deep learning*) para o desenvolvimento de um sistema

classificador binário de imagens de raio-x pulmonares, com o objetivo de identificar o modelo mais eficaz para distinguir pulmões saudáveis daqueles afetados por pneumonia, sem traçar a distinção entre a viral e a bacteriana.

II. REVISÃO DE LITERATURA

Hasan, Alom e Ali [7] exploraram o uso de modelos de *deep learning* para identificar casos de COVID-19 e pneumonia. Utilizando a base de dados *chestX-ray-dataset*, composta por 701 imagens de COVID-19 e 5863 de pneumonia, os autores testaram e compararam os modelos *EfficientNet*, VGG e *DenseNet*. A implementação foi realizada em Python com Keras, e o modelo *DenseNet* destacou-se com uma acurácia média de 92%, demonstrando seu potencial para a classificação de imagens médicas.

O estudo de Ouerhani, Boulares e Mahjoubi [12] focou na detecção de pneumonia pediátrica através de imagens de raio-x torácicas. Utilizando o banco de dados de Guangzhou, composto de 5863 imagens de pneumonia, os autores compararam os modelos *Inception*, VGG e *ResNet* em um ambiente de Jupyter notebook no Kaggle. O modelo VGG apresentou a melhor acurácia, 92%, e os autores ressaltam a utilidade dessa tecnologia no suporte ao diagnóstico médico, minimizando subjetividades e erros humanos.

Sharma *et al.* [11] apresentaram dois modelos de CNN para a identificação de pneumonia em radiografias pulmonares, diferenciados pelo uso de *dropout*. Ambos incluíram camadas de convolução, *max pooling* e classificação, e foram treinados com a base de dados de Guangzhou. A adição de técnicas de *data augmentation* e *dropout* resultou em uma acurácia de 90,68%, sugerindo que essas abordagens podem aumentar a robustez do modelo. Os autores indicam o potencial para futuros aprimoramentos, como *early stopping* e *batch normalization*.

Stephen *et al.* [13], por sua vez, desenvolveram um modelo de CNN sem recorrer ao *transfer learning*, utilizando a base de dados de Guangzhou. Com o auxílio da técnica de *data augmentation*, o modelo personalizado alcançou uma acurácia de 95,31%, destacando a viabilidade de modelos desenvolvidos do zero.

Em um estudo comparativo, Ikechukwu *et al.* [8] avaliaram os desempenhos dos modelos VGG e *ResNet* para a detecção de pneumonia, também utilizando *data augmentation* e *dropout*. Com a base de dados de Guangzhou, o modelo VGG superou o ResNet, alcançando uma acurácia de 97,3% contra 96,2%.

Ferreira [6] desenvolveu um método para a detecção de pneumonia em imagens de raio-x usando a base de dados de Guangzhou. A etapa de pré-processamento utilizada envolveu a remoção de regiões anatômicas não relevantes para o treinamento, além da equalização, para aumentar o contraste e melhorar o reconhecimento de padrões. O modelo que obteve os melhores resultados foi o VGG, alcançando uma acurácia de 97,4%.

Finalmente, Alsharif *et al.* [1] propuseram uma CNN simplificada, com menos camadas que modelos tradicionais como

ResNet e *DenseNet*. Utilizando a base de dados de Guangzhou, além de técnicas como *max pooling* e *batch normalization*, eles obtiveram uma acurácia impressionante de 99,72%, mostrando que uma arquitetura mais simples pode ainda ser altamente eficaz.

As pesquisas revisadas demonstram o progresso significativo no uso de modelos de *deep learning* para a detecção de pneumonia em radiografias. Modelos como VGG e *ResNet* são frequentemente destacados, e técnicas como *data augmentation*, *dropout* e *batch normalization* têm se mostrado eficazes na melhoria da precisão diagnóstica. Outro ponto a se destacar é a utilização da base de Guangzhou com imagens de pulmões saudáveis e comprometidos com pneumonia em cinco dos trabalhos revisados, o que permite uma comparação interessante entre a literatura e os resultados obtidos neste trabalho. [1], [6], [8], [11]–[13]

III. MATERIAIS E MÉTODOS

O desenvolvimento do classificador foi realizado no ambiente Google Colab, utilizando o *backend* do L4 GPU para lidar com o alto consumo de RAM durante as análises.

Inicialmente, a base de dados de Guangzhou foi analisada, revelando um desbalanceamento significativo entre o número de imagens de pulmões normais (1583) e de pulmões afetados por pneumonia bacteriana ou viral (4273). Para mitigar esse problema e reduzir o risco de *overfitting*, duas estratégias foram adotadas: primeiramente, aplicou-se a técnica de *data augmentation* para gerar 427 novas imagens, representando 27% da classe minoritária, conforme sugerido por Buda, Maki e Mazurowski [2]. Além disso, utilizaram-se pesos de classe durante o ajuste dos modelos para atenuar o desbalanceamento.



Figura 1. Exemplo de imagens presentes na base de Guangzhou

Após a expansão da base com as novas imagens, as pastas de treinamento e teste foram combinadas e pré-processadas para imagens em escala de cinza. As imagens foram redimensionadas para 224x224 pixels para os modelos tradicionais. Para os modelos de aprendizado profundo, as imagens de entrada apresentam três canais RGB, portanto as imagens foram redimensionadas para 224x224x3 pixels, mas se mantiveram em escala de cinza, de acordo com as recomendações da documentação do Keras [9].

Em seguida, as imagens foram divididas em conjuntos de treinamento e teste usando o método *train_test_split*, com uma proporção de 70/30 e *random state* de 35. Foram criadas duas *pipelines*: uma para modelos tradicionais, incluindo Regressão Logística, Árvore de Decisão, Máquina de Vetores de Suporte (SVM), *Random Forest* e uma Rede Neural Multicamada

(MLP) de cinco camadas densas; e outra para modelos de aprendizado profundo, como VGG16, *ResNet50* e *Efficient-NetB0*.

Na primeira *pipeline*, utilizou-se *cross validation* para avaliar o *F1 score* de cada modelo. Essa métrica foi empregada para comparar os modelos e serviu como critério principal para selecionar o melhor classificador entre os modelos tradicionais.

Na segunda *pipeline*, foram aplicadas técnicas de *dropout* (com taxa de 0,2) e *early stopping*. O *F1 score* também foi utilizado como critério para escolher o melhor modelo. Após a seleção dos melhores modelos de cada *pipeline*, suas métricas e matrizes de confusão foram comparadas para determinar o modelo mais adequado para a tarefa de classificação.

Durante a análise, optou-se por não utilizar *batch normalization*, pois as imagens já passaram por um pré-processamento e normalização, e o volume de dados não justificava a aplicação dessa técnica.

IV. RESULTADOS

Os resultados obtidos a partir dos experimentos forneceram importantes *insights* sobre a eficácia e desempenho dos diferentes modelos para classificação de imagens de raio-x pulmonares.

A. Modelos baseados em Aprendizado de Máquina Tradicionais

Foram avaliados cinco modelos tradicionais de aprendizado de máquina: Regressão Logística, Árvore de Decisão, SVM, *Random Forest* e MLP.

- **Regressão Logística:** Este modelo alcançou uma pontuação média de *F1 score* de 0,92, acurácia de 0,90, precisão de 0,91 e *recall* de 0,95, demonstrando um desempenho consistente.
- **Árvore de Decisão:** Obteve uma média de *F1 score* de 0,89, acurácia de 0,87, precisão de 0,90 e *recall* de 0,92, o que representa um desempenho ligeiramente inferior em comparação com os outros modelos.
- **SVM:** Destacou-se com uma pontuação média de *F1 score* de 0,96, acurácia de 0,95, precisão de 0,96 e *recall* de 0,96, superando os demais modelos testados.
- **Random Forest:** Este modelo de conjunto apresentou uma média de *F1 score* de 0,95, acurácia de 0,93, precisão de 0,93 e *recall* de 0,97, ficando próximo ao desempenho da SVM, mas um pouco abaixo.
- **MLP de cinco camadas:** Obteve uma média de *F1 score* de 0,94, acurácia de 0,92, precisão de 0,93 e *recall* de 0,95, demonstrando um desempenho competitivo.

Os valores de *F1 score* para os modelos baseados em aprendizado de máquina tradicional estão apresentados na Tabela 1.

Com os resultados obtidos dos cinco modelos testados, aquele que obteve os melhores resultados pela métrica de avaliação usada foi o SVM. A Matriz de Confusão do modelo SVM pode ser vista na Figura 1, onde o rótulo 0 corresponde à classe saudável e 1 corresponde à classe pneumonia.

Tabela I
RESULTADOS DOS MODELOS BASEADOS EM APRENDIZADO DE MÁQUINA TRADICIONAL

Modelo de Aprendizado de Máquina Tradicional	<i>F1 Score</i>
Regressão Logística	0,92
Árvore de Decisão	0,89
SVM	0,96
<i>Random Forest</i>	0,95
MLP	0,94

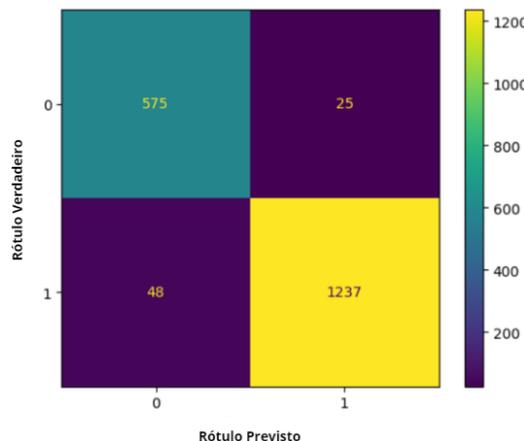


Figura 2. Matriz de Confusão - SVM.

B. Modelos baseados em Aprendizado Profundo

Para os modelos de aprendizado profundo, foram testadas e comparadas três abordagens distintas: Redes Neurais Convolucionais (CNNs) dos tipos VGG, *EfficientNet* e *ResNet*.

- **VGG:** Este modelo alcançou uma média de *F1 score* de 0,96, acurácia de 0,94, precisão de 0,95 e *recall* de 0,96, confirmando um bom desempenho na classificação de imagens.
- **EfficientNet:** Apresentou uma pontuação média de *F1 score* de 0,98, acurácia de 0,97, precisão de 0,97 e *recall* de 0,98, demonstrando uma excelente capacidade de classificação.
- **ResNet:** Também alcançou uma média de *F1 score* de 0,98, acurácia de 0,97, precisão de 0,99, e *recall* de 0,97, equiparando-se ao *F1 score* da *EfficientNet* e ligeiramente superando em termos de classificações corretas.

Tabela II
RESULTADOS DOS MODELOS BASEADOS EM APRENDIZADO PROFUNDO

Modelo de Aprendizado Profundo	<i>F1 Score</i>
VGG	0,96
ResNet	0,98
EfficientNet	0,98

Com os resultados dos três modelos de aprendizagem profunda, a *ResNet* foi a que obteve os resultados mais significativos de acordo com a métrica utilizada. A Matriz de Confusão do modelo *ResNet* pode ser vista na Figura 2.

Os resultados dos modelos baseados em aprendizado profundo estão sumarizados na Tabela 2.

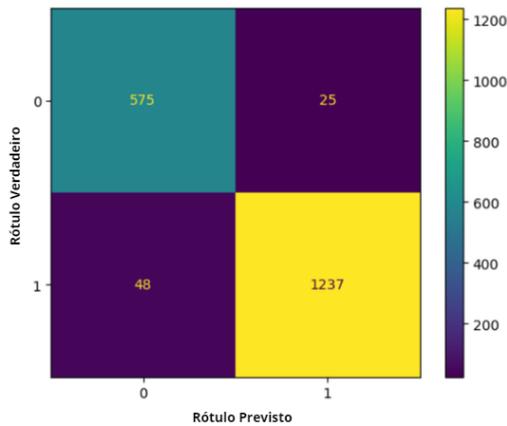


Figura 3. Matriz de Confusão - ResNet.

V. CONCLUSÃO

Este estudo realizou uma comparação de diferentes estratégias para a classificação de imagens de raio-x pulmonares, utilizando tanto modelos de aprendizado de máquina tradicionais quanto baseados em aprendizado profundo. Os testes mostraram resultados promissores na tarefa de classificação, empregando uma base de dados que foi cuidadosamente pré-processada, incluindo normalização e balanceamento de classes.

Os experimentos indicaram que os modelos de aprendizado profundo, especialmente as redes neurais convolucionais como *ResNet* e a *EfficientNet*, superaram significativamente os modelos tradicionais em termos de desempenho. As pontuações de *F1 score* desses modelos destacaram sua maior capacidade de generalização e habilidade em identificar características relevantes nas imagens de raio-x.

A arquitetura *ResNet* obteve os melhores resultados. Isso é devido às suas conexões de atalho que ajudam a enfrentar os desafios de treinar redes profundas, melhorando a interpretabilidade, o desempenho e capacidade de extrair características significativas, contribuindo para a análise de imagens médicas [14]. Esses fatores podem indicar a *ResNet* como uma escolha assertiva para a classificação de imagens de raio-x em comparação com outras abordagens.

O uso de aprendizado profundo para a classificação de imagens tem um enorme potencial para auxiliar profissionais de saúde, permitindo a identificação mais eficiente de problemas pulmonares. Diagnósticos mais rápidos e precisos possibilitam o início de tratamentos de forma mais precoce, evitando complicações graves e otimizando o uso de recursos na área da saúde.

A qualidade das imagens é crucial para a eficácia das redes neurais convolucionais, como a *ResNet*, em razão do princípio da construção direta de características a partir de imagens, que é a base para esse modelo de arquitetura. Em decorrência disso, essas redes são exigentes em relação à qualidade da imagem e à homogeneidade das amostras de treinamento e validação [5]. Com o uso dessa abordagem, é possível fornecer suporte significativo aos profissionais na

realização de diagnósticos precisos e na implementação de tratamentos mais eficazes.

REFERÊNCIAS

- [1] Alsharif, Roaa and Al-Issa, Yazan and Alqudah, Ali Mohammad and Qasmieh, Isam Abu and Mustafa, Wan Azani and Alquran, Hiam, "PneumoniaNet: Automated Detection and Classification of Pediatric Pneumonia Using Chest X-ray Images and CNN Approach". *Electronics*, v. 10, n. 23, p. 2949, 2021.
- [2] Buda, Mateusz and Maki, Atsuto and Mazurowski, Maciej A, "A systematic study of the class imbalance problem in convolutional neural networks". *Neural networks*, v. 106, p. 249-259, 2018. Elsevier.
- [3] Chee, Elyssa and Huang, Kathryn and Haggie, Stuart and Britton, Philip N, "Systematic review of clinical practice guidelines on the management of community acquired pneumonia in children". *Paediatric Respiratory Reviews*, v. 42, p. 59-68, 2022. Elsevier.
- [4] Corrêa, Ricardo de Amorim and Costa, Andre Nathan and Lundgren, Fernando and Michelin, Lessandra and Figueiredo, Mara Rúbia and Holanda, Marcelo and Gomes, Mauro and Teixeira, Paulo José Zimmermann and Martins, Ricardo and Silva, Rodney and others, "2018 recommendations for the management of community acquired pneumonia". *Jornal Brasileiro de Pneumologia*, v. 44, n. 05, p. 405-423, 2018. SciELO Brasil.
- [5] Dovganich, A. A. and Khvostikov, A. V. and Krylov, A. S. and Parolina, L. E., "Automatic quality control in lung X-ray imaging with deep learning". *Computational Mathematics and Modeling*, v. 32, p. 276-285, 2021.
- [6] Ferreira, José Raniery and Cardenas, Diego Armando Cardona and Moreno, Ramon Alfredo and de Sá Rebelo, Marina de Fátima and Krieger, José Eduardo and Gutierrez, Marco Antonio, "Multi-view ensemble convolutional neural network to improve classification of pneumonia in low contrast chest x-ray images". In: 2020 42nd annual international conference of the IEEE engineering in Medicine & Biology Society (EMBC). IEEE, p. 1238-1241, 2020.
- [7] Hasan, Md Jahid and Alom, Md Shahn and Ali, Md Shikhar, "Deep learning based detection and segmentation of COVID-19 & pneumonia on chest X-ray image". In: 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD). IEEE, p. 210-214, 2021.
- [8] Ikechukwu, A Victor and Murali, S and Deepu, R and Shivamurthy, RC, "ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images". *Global Transitions Proceedings*, v. 2, n. 2, p. 375-381, 2021. Elsevier.
- [9] Keras, "Keras 3 API documentation: Keras Applications". 2024. <https://keras.io/api/applications/vgg/>. Acessado em 13/07/2024.
- [10] Mujahid, Muhammad and Rustam, Furqan and Álvarez, Roberto and Luis Vidal Mazón, Juan and Díez, Isabel de la Torre and Ashraf, Imran, "Pneumonia classification from X-ray images with inception-V3 and convolutional neural network". *Diagnostics*, v. 12, n. 5, p. 1280, 2022. MDPI.
- [11] Sharma, Harsh and Jain, Jai Sethia and Bansal, Priti and Gupta, Sumit, "Feature extraction and classification of chest x-ray images using cnn to detect pneumonia". In: 2020 10th international conference on cloud computing, data science & engineering (Confluence). IEEE, p. 227-231, 2020.
- [12] Ouerhani, Amira and Boulares, Souhaila and Mahjoubi, Halima, "Automated Detection of Pediatric Pneumonia from Chest X-Ray Images Using Deep Learning Models". In: 2023 IEEE Afro-Mediterranean Conference on Artificial Intelligence (AMCAI). IEEE, p. 1-7, 2023.
- [13] Stephen, Okeke and Sain, Mangal and Maduh, Uchenna Joseph and Jeong, Do-Un, "An efficient deep learning approach to pneumonia classification in healthcare". *Journal of healthcare engineering*, v. 2019, n. 1, p. 4180949, 2019. Wiley Online Library.
- [14] Xu, Wannan and Fu, You-Lei and Zhu, Dogmei, "ResNet and its application to medical image processing: Research progress and challenges". *Computer Methods and Programs in Biomedicine*, v. 240, p. 107660, 2023.
- [15] Yun, Ki Wook and Wallihan, Rebecca and Juergensen, Alexis and Mejias, Asuncion and Ramilo, Octavio, "Community-acquired pneumonia in children: myths and facts". *American journal of perinatology*, v. 36, n. S 02, p. S54-S57, 2019. Thieme Medical Publishers.