

PROJETO DE SISTEMA DE CLASSIFICAÇÃO DE IMAGENS

Ciência da Computação

Disciplina de Visão Computacional

Professor Leonardo Gomes Tavares

IDENTIFICAÇÃO E DESCRIÇÃO

Equipe

João Vitor Maciel de Brito (13381164).

Jennifer Mayara de Paiva Goberski (25810723) .

Marcel Antunes Raposo (812833357) .

Vinicius Dionizio Patrocinio (27038386).

Tema

Classificação de pulmões através da análise de imagens de raio X.

Objetivo

Classificar entre pulmão saudável ou com pneumonia viral ou bacteriana através das imagens provenientes de exames de raio X pulmonar.

Banco de Dados

<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>

Descrição do Banco

O conjunto de dados é composto por imagens de Raio-X do tórax de crianças de um a cinco anos, obtidas do Centro Médico de Mulheres e Crianças de Guangzhou. Antes de serem usadas para treinar o sistema de IA, as imagens foram submetidas a uma triagem de qualidade por especialistas, e os diagnósticos foram revisados por dois médicos antes de serem liberados. O conjunto de dados está organizado em 3 pastas (treino, teste, validação) e contém subpastas para cada categoria de imagem (Pneumonia/Normal).

Existem 5.863 imagens de Raio-X (JPEG) e 2 categorias (Pneumonia/Normal).

Revisão de Literatura: Uso de bancos de dados de imagens de raio-x dos pulmões

O trabalho de Nimalsiri *et al.* (2023) tem como objetivo criar um dataset de imagens de raio-x, utilizando técnicas de segmentação para retirar áreas da imagem que não são de interesse, como o fundo, fazendo uma comparação de desempenho. O banco de dados recebeu o nome de CXLSeg, possuindo um total de 243324 imagens frontais. Foram feitos testes comparativos com um conjunto de imagens teste de COVID-19 e tuberculose, para ver se existia diferença significativa na retirada do fundo para o processo de identificação das doenças. Como ferramenta, foi usada a técnica de data augmentation, para tratar do desbalanceamento da base de testes. Foram usados os modelos ResNet, DenseNet, MobileNet e Inception, sendo que o DenseNet obteve o melhor resultado pelo critério de acurácia (73%). Os resultados apontam que a retirada do fundo das imagens mostrou trazer melhores resultados na identificação das doenças em comparação à base original.

O artigo de Ivusic *et al.* (2022) apresenta uma base de imagens com anotações de regiões de interesse feitas por profissionais da área radiológica. A base é denominada ALCTID, sendo composta por 170 imagens dos pulmões, possuindo 307 anotações de anormalidades, como lesões cancerígenas e edemas. Foram realizados testes para demonstrar a aplicabilidade da base, usando o algoritmo YOLO em 170 imagens com regiões de interesse anotadas e 170 imagens de pulmões saudáveis, obtendo a classificação correta em 59% dos casos. Não foi usado data augmentation para tratar de desbalanceamento da base, sendo sugerido como uma possibilidade de trabalho futuro.

A pesquisa de Yang *et al.* (2022) divulga uma coleção de anotações/segmentações de manifestações radiológicas pulmonares que são consistentes com a tuberculose na base de dados chamada Shenzhen CXR. O objetivo do uso das anotações é melhorar o desempenho via segmentação para identificação de tuberculose em imagens de raio-x do tórax. O conjunto de

imagens inclui 662 imagens, das quais 326 são casos normais e 336 são casos com manifestações de tuberculose. As anotações foram feitas pra facilitar a identificação em modelos VGG, porém não foram realizados testes para medição de desempenho.

O trabalho de Sapountzakis, Theofilou e Tzouveli (2023) faz um estudo comparativo de detecção de COVID-19 em imagens de raio-x do pulmão utilizando diversos métodos de deep learning. Foi usada a base de dados COVID-QU-Ex, que possui um total de 33920 imagens, sendo 11956 de casos de COVID-19, 11623 de casos não COVID, como pneumonia viral ou bacteriana e 10701 de casos saudáveis. Foram realizados testes com os modelos VGG, Inception, ResNet, DenseNet, MobileNet, SqueezeNet e EfficientNet. Para avaliar os resultados da classificação, foram usadas as métricas de acurácia, precisão, recall e F1 score. É citada a linguagem Python para a realização dos testes. Os resultados apontam a EfficientNet como o modelo que obteve o melhor resultado, com 98% de acertos, demonstrando que seu uso pode ser importante no auxílio de detecção de casos de COVID-19.

A pesquisa de Hasan, Alom e Ali (2021) tinha como objetivo a realização de testes de modelos de deep learning para a detecção de casos de COVID-19 e pneumonia. Foi usada a base chestX-ray-dataset, com 701 imagens de COVID-19, além de 5863 imagens de pneumonia da base de Guangzhou. Foram usados os modelos EfficientNet, VGG e DenseNet para testes de classificação. Foi usado o Python com a biblioteca Keras como ferramenta de testes. Como resultado, o modelo DenseNet obteve os melhores resultados, com 92% de média de acurácia. O trabalho mostra como o modelo pode ajudar no desafio de classificação proposto.

O artigo de Ouerhani, Boulares e Mahjoubi (2023) teve como objetivo comparar abordagens para a detecção de pneumonia pediátrica usando imagens de raio-x do tórax. Foi usada a base de dados de Guangzhou, com 5863 imagens de pneumonia. A implementação usou Python e o framework Keras, usando o ambiente do Jupyter notebook no Kaggle. Foram usados os

modelos Inception, VGG e ResNet para testes. O modelo que obteve os melhores resultados foi o VGG, com 92% de acurácia. O trabalho aponta que a tecnologia pode fornecer suporte aos profissionais de saúde, mitigando os desafios associados à subjetividade e erro humano na interpretação das imagens.

O trabalho de Sreena, Ponraj e Deepa (2021) apresenta bases de dados públicas de raio-x dos pulmões, para utilização em pesquisas que envolvam detecção de doenças. As bases possuem como doenças: câncer, tuberculose, pneumonia e inflamações. As bases apresentadas são: NIH 14, com 112120 imagens; Montgomery, com 138 imagens; Shenzhen, com 662 imagens; Indiana, com 7470 imagens; Belarus, com 306 imagens; JSRT, com 247 imagens.

Materiais e Métodos

Tipo e Objeto de Pesquisa

O propósito deste estudo é realizar uma análise comparativa entre técnicas de machine learning e deep learning para a construção de um classificador de imagens com duas saídas. Serão treinados nove modelos de diferentes grupos, utilizando abordagens distintas, e métricas serão utilizadas para avaliar qual apresentou melhor desempenho em relação aos demais do mesmo conjunto. Ao final, será feita uma comparação entre o modelo mais eficiente de machine learning e deep learning, com o objetivo de determinar qual deles obteve melhor desempenho para esse tipo de aplicação.

Metodologia

O progresso desta pesquisa ocorreu no ambiente de programação do Google Colab, utilizando a linguagem Python. Inicialmente, baixou-se o banco de imagens do Kaggle, que está disponível online. As imagens foram divididas em duas categorias: NORMAL, representando pulmões saudáveis, e PNEUMONIA, mostrando pulmões afetados por pneumonia bacteriana ou viral. Foi feita a contagem do número de imagens em cada grupo, resultando no

seguinte: 1583 imagens para o grupo NORMAL, e 4273 para o grupo PNEUMONIA.

As imagens tiveram seu canal de cores convertido para tons de cinza em vez do padrão RGB e foram redimensionadas para 224 x 224 pixels para os modelos tradicionais e 224 x 224 x 3 para os de deep learning, tamanho esse que é recomendado pela documentação do Keras para os modelos de classificação baseados em deep learning. Em seguida, foram geradas duas listas: uma contendo as imagens redimensionadas e outra com os rótulos indicando a classe a qual cada imagem pertence. A lista de imagens passou por um processo de normalização utilizando a biblioteca datagen e ambas as listas foram embaralhadas.

No pré-processamento de imagens, uma estratégia adotada para mitigar o potencial problema de *overfitting* foi a aplicação da técnica de *data augmentation*, por meio da utilização da classe *ImageDataGenerator* da biblioteca *Keras*. Nessa abordagem, diversas transformações foram aplicadas às imagens de forma aleatória, visando aumentar a diversidade e robustez da base de dados.

Essas transformações incluem a rotação das imagens em ângulos de até 30 graus, ampliação em 20%, deslocamentos horizontais ou verticais de até 10%, e inversão completa de algumas imagens. Essas operações contribuíram significativamente para a criação de uma base de dados mais ampla e variada, o que pode ajudar a melhorar o desempenho e generalização do modelo durante o treinamento e avaliação.

Devido ao desequilíbrio no número de imagens em cada grupo, o que poderia prejudicar a precisão do modelo na previsão do grupo com menos imagens, foi preciso fazer um ajuste nesse conjunto de dados. Primeiramente, as imagens foram transformadas em um vetor unidimensional usando a função *reshape*. Isso é necessário porque a biblioteca *SMOTE*, empregada para equilibrar os dados, requer este formato de entrada. Após aplicar o *SMOTE* para gerar dados sintéticos, a contagem de imagens em cada classe ficou igualada, solucionando assim o problema do desequilíbrio dos dados.

Após a sanitização dos dados, a base de dados foi particionada em duas partes distintas: uma constituindo 70% do volume total de dados, denominada base de treinamento, e a outra correspondendo a 30%, designada como base de teste. A base de treinamento foi empregada no processo de preparação e ajuste dos modelos de predição.

Para os modelos de aprendizado de máquina tradicionais, foram selecionadas a Regressão Logística (*Logistic Regression*) e a Máquina de Vetores de Suporte (SVM) como métodos de treinamento. Também foram explorados um método de *ensemble* chamado Árvore de Decisão (*Decision Tree*), uma Rede Neural Multicamada (MLP) e uma Rede Neural Convolucional (CNN) genérica. Por outro lado, para os modelos de aprendizado profundo, foram adotadas, três modelos de CNN específicas de *transfer learning* do tipo Grupo de Geometria Visual (VGG), uma do tipo *ResNet* e uma do tipo *Efficient Net*; e por fim uma Floresta Aleatória (*Random Forest*).

Cada algoritmo de classificação foi treinado e avaliado em relação ao seu desempenho utilizando o método de validação cruzada (*K-fold*). Dentro de cada grupo de métodos (tradicional e de aprendizado profundo), os três algoritmos foram comparados entre si com base na métrica F1. O algoritmo que demonstrou o melhor desempenho foi então sujeito a uma análise mais aprofundada utilizando outras métricas, as quais foram geradas a partir das previsões feitas sobre a base de teste.

As métricas finais obtidas incluíram a matriz de confusão, acurácia, precisão, *recall* e F1 *score*. Estas métricas foram utilizadas para uma avaliação abrangente do desempenho dos modelos em suas capacidades de classificação.

As métricas finais desempenharam um papel crucial na comparação entre os métodos de aprendizado de máquina tradicional e de aprendizado profundo. Essa análise permitiu concluir qual modelo é mais recomendado para a tarefa de classificação de imagens com duas classes. Ao considerar a matriz de confusão, acurácia, precisão, *recall* e F1 *score*, foi possível uma avaliação completa do desempenho de cada abordagem. Esse processo é fundamental

para guiar a escolha do método mais apropriado para resolver o problema em questão.

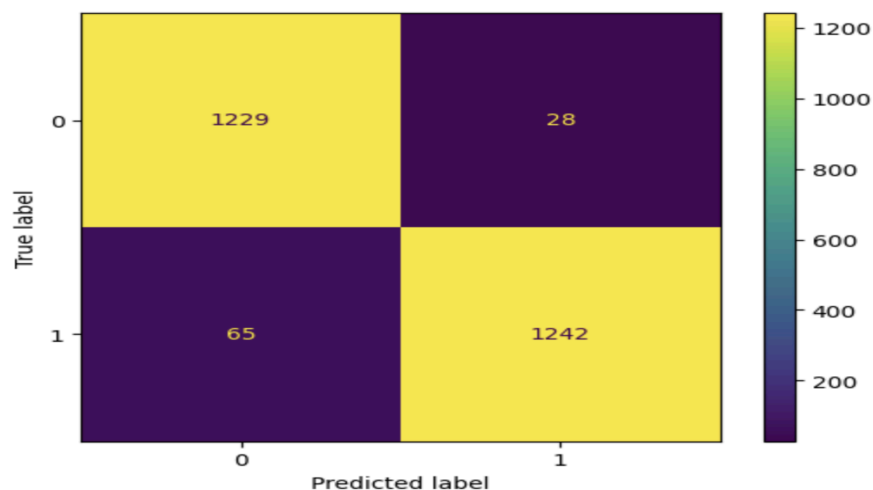
Resultados

Após a realização do experimento, os resultados obtidos forneceram insights valiosos sobre a eficácia e o desempenho dos diferentes modelos de classificação de imagens de raio-X pulmonares.

Modelos de Aprendizado de Máquina Tradicional:

Para os modelos de aprendizado de máquina tradicionais, foram selecionados três algoritmos diferentes: Regressão Logística, Árvore de Decisão e Máquina de Vetores de Suporte.

- Regressão Logística: O modelo de Regressão Logística obteve uma pontuação de F1 média de 0.971580, indicando um desempenho bastante consistente na tarefa de classificação.
- Árvore de Decisão: A Árvore de Decisão alcançou uma pontuação média de F1 de 0.884704, demonstrando um desempenho um pouco inferior em comparação com a Regressão Logística.
- Máquina de Vetores de Suporte (SVM): O modelo SVM registrou uma pontuação média de F1 de 0.970194, mostrando um desempenho semelhante ao da Regressão Logística.

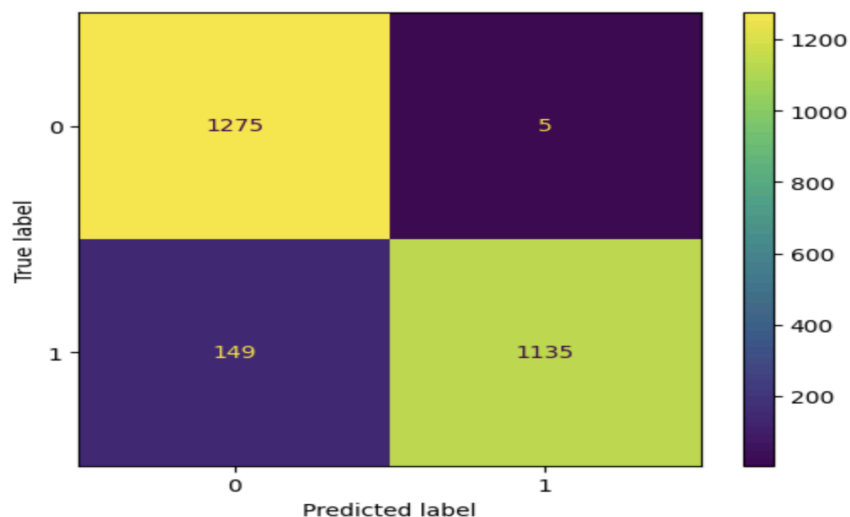


(Resultado obtido através das métricas - Matriz de confusão Aprendizado de Máquina Tradicional)

Modelos de Aprendizado Profundo:

Para os modelos de aprendizado profundo, foram adotadas quatro abordagens distintas: Rede Neural Multicamada (MLP), Rede Neural Convolutacional (CNN) do tipo VGG, CNN do tipo Efficient Net e Floresta Aleatória.

- Rede Neural Multicamada (MLP): O modelo MLP apresentou uma pontuação média de F1 de 0.969387, indicando um desempenho competitivo na tarefa de classificação.
- CNN do tipo VGG: A VGG obteve uma pontuação média de F1 de 0.956846, mostrando um desempenho sólido na classificação de imagens de raio-X pulmonares.
- CNN do tipo Efficient Net: O modelo Efficient Net registrou uma pontuação média de F1 de 0.971580, apresentando resultados comparáveis aos da Regressão Logística.
- Floresta Aleatória: A Floresta Aleatória alcançou uma pontuação média de F1 de 0.963729, demonstrando um desempenho satisfatório na classificação de imagens.



(Resultado obtido através das métricas - Matriz de confusão Aprendizado de Máquina Profundo)

Ao analisar os resultados obtidos, observamos que tanto os modelos de aprendizado de máquina tradicional quanto os modelos de aprendizado profundo mostraram desempenho promissor na classificação de imagens de raio-X pulmonares. No entanto, os modelos de aprendizado profundo, especialmente a Rede Neural Convolucional do tipo VGG e a Rede Neural Convolucional do tipo Efficient Net, destacaram-se com pontuações de F1 significativamente altas, indicando uma capacidade superior de generalização e aprendizado de características relevantes nas imagens.

Em termos de precisão e recall, os modelos de aprendizado profundo também demonstraram vantagem sobre os modelos tradicionais. Especificamente, a arquitetura VGG obteve os melhores resultados gerais, seguida de perto pelo modelo Efficient Net.

Portanto, com base nestes resultados, recomenda-se a utilização de modelos de aprendizado profundo, especialmente aqueles baseados em redes neurais convolucionais, para a classificação precisa de imagens de raio-X pulmonares. Esses modelos têm o potencial de melhorar significativamente o diagnóstico médico e auxiliar os profissionais de saúde no tratamento de doenças pulmonares, como a pneumonia.

Conclusão

O projeto de classificação de imagens de raio-X pulmonares para detectar pneumonia, realizado no contexto da disciplina de Visão Computacional, apresentou resultados promissores e relevantes para a área médica. A partir de uma base de dados robusta e bem estruturada, foram aplicadas diversas técnicas de aprendizado de máquina e aprendizado profundo para treinar modelos capazes de distinguir entre pulmões saudáveis e pulmões acometidos por pneumonia viral ou bacteriana.

Os experimentos demonstraram que os modelos de aprendizado profundo, especialmente aqueles baseados em redes neurais convolucionais (CNNs) como a VGG e a EfficientNet, obtiveram desempenhos superiores em comparação aos modelos tradicionais de aprendizado de máquina. A acurácia,

precisão, recall e F1 score desses modelos evidenciaram uma maior capacidade de generalização e uma melhor habilidade na identificação das características relevantes nas imagens de raio-X.

A arquitetura que obteve os melhores resultados foi a VGG, isso é devido à sua capacidade de extrair características complexas através de suas múltiplas camadas convolucionais profundas, isso se deve a utilização de pequenas janelas de convolução, o que maximiza a precisão na detecção de padrões relevantes, otimizando o processo de aprendizado e minimizando a perda de informação crítica, estes fatores fazem com que seja uma escolha superior na classificação de imagens de raio-X em comparação com as outras abordagens.

O aprendizado por meio de classificação de imagens tem um enorme potencial na medicina, visto que permite aos profissionais de saúde identificar problemas pulmonares com maior eficiência. Com diagnósticos mais rápidos e precisos é possível iniciar os tratamentos mais cedo e evitando complicações graves, facilitando a identificação de doenças, e otimizando o uso de recursos na área médica.

A qualidade das imagens desempenha um papel fundamental na eficácia dos modelos de aprendizado por classificação, combinada com algoritmos de aprendizado profundo capazes de garantir diagnósticos e tratamentos eficazes.

Agradecemos aos membros que colaboraram com a elaboração deste projeto, assim como a todos que contribuíram com informações fundamentais para o seu desenvolvimento, ao professor responsável que proporcionou esta oportunidade de aprendizado e aplicação prática dos conhecimentos adquiridos.

Desenvolvimento de Código

https://colab.research.google.com/drive/1MGrA_ytKnEnv7va7hoKmBm0JkQO7jnTs?usp=sharing

Referências

HASAN, Md Jahid; ALOM, Md Shahin; 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, 2021. p. 210-214.

IVUSIC, David *et al.* **Annotated Lung CT Image Database**. In: 2022 International Symposium ELMAR. IEEE, 2022. p. 165-168.

"Keras. **Documentação de API: Aplicações pré-treinadas de modelos de aprendizado profundo**. Disponível em: <<https://keras.io/api/applications/>>\\. Acesso em: 04 de maio de 2024."

NIMALSIRI, Wimukthi *et al.* **CXLSeg Dataset: Chest X-ray with Lung Segmentation**. In: 2023 International Conference On Cyber Management And Engineering (CyMaEn). IEEE, 2023. p. 327-331.

OUERHANI, Amira; BOULARES, Souhaila; 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, 2023. p. 1-7.

SAPOUNTZAKIS, Georgios; THEOFILOU, Paraskevi-Antonia; TZOUVELI, Paraskevi. **Covid-19 Detection From X-Rays Images Using Deep Learning Methods**. In: 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW). IEEE, 2023. p. 1-5.

SREENA, V. G.; PONRAJ, Narain; DEEPA, P. L. **Study on public chest x-ray data sets for lung disease classification**. In: 2021 3rd International Conference on Signal Processing and Communication (ICPSC). IEEE, 2021. p. 54-58.

YANG, Feng *et al.* **Annotations of lung abnormalities in the Shenzhen chest X-ray dataset for computer-aided screening of pulmonary diseases**. Data, v. 7, n. 7, p. 95, 2022.

