

Modelo híbrido de aprendizado de máquina para extração de características em imagens de radiografia de tórax

Rodrigo Marcel Araujo Oliveira
PEI - Escola Politécnica
Universidade Federal da Bahia
Salvador, Bahia
email:rodrigomarcel@ufba.br

Caroline Ribeiro Figueredo
PPGEE - Escola Politécnica
Universidade Federal da Bahia
Salvador, Bahia
email: carolinefigueredo@ufba.br

Luciano Almeida da Silva
PPGEE - Escola Politécnica
Universidade Federal da Bahia
Salvador, Bahia
email:lucianoas@ufba.br

Paulo César Machado de Abreu Farias
PPGEE - Escola Politécnica
Universidade Federal da Bahia
Salvador, Bahia
email:paulo.farias@ufba.br

Ângelo Márcio Oliveira Sant'Anna
PEI - Escola Politécnica
Universidade Federal da Bahia
Salvador, Bahia
email: angelo.santanna@ufba.br

Resumo—O diagnóstico de doenças com imagens e auxílio de sistemas computacionais de Inteligência Artificial vêm sendo desenvolvidos e aplicados em vários contextos clínicos. A Indústria 4.0 tem um papel fundamental nesse processo e pesquisas recentes demonstram que transferência de aprendizado de redes neurais convolucionais pré-treinadas como ResNet-50, VGG-16 e Inception podem contribuir significativamente no resultado das predições para classificação de imagens. Contudo, na maioria das vezes, o treinamento de redes neurais requer um tempo de execução considerável. Este artigo propõe um método rápido e eficiente para treinamento com menor tempo de execução baseado em algoritmos de aprendizado de máquina para modelagem preditiva de imagens de radiografia de tórax de pacientes saudáveis e com pneumonia. O trabalho consiste em desenvolver uma modelagem híbrida usando a ResNet-50 para extração de características, a análise de componentes principais para reduzir a dimensão das variáveis preditoras e o modelo *Random Forest* para classificar as imagens. Algoritmos genéticos foram utilizados para otimizar os parâmetros do modelo *Random Forest*. Os resultados dessa abordagem permitiu alcançar valores relevantes de desempenho quando comparado com os encontrados na literatura por arquiteturas robustas de redes neurais. O classificador híbrido *Random Forest* proposto alcançou 93%, 79%, 88% de ROC-AUC nos conjuntos de treino, teste e validação, respectivamente.

Palavras-chave—ResNet-50, *Random Forest*, radiografia de tórax, Algoritmos Genéticos, PCA.

I. INTRODUÇÃO

A pneumonia é uma doença infecciosa que se instala nos pulmões, no mundo aproximadamente 1 milhão de crianças morrem de pneumonia, ela é a principal causa de morte de crianças menores de cinco anos [1]. No Brasil, há 4 milhões de casos de pneumonia infantil a cada ano, em que 11% das mortes são em crianças abaixo de um ano e 13% entre crianças de 1 a 4 anos. As taxas de mortalidade infantil por pneumonias

variam por região, sendo mais altas nos estados do Norte e Nordeste e mais baixas no Sul [2] [3] [4].

O diagnóstico de pneumonia em crianças pode ser feito a partir do histórico do paciente e do exame clínico, podendo ser confirmado com uma radiografia de tórax. A radiografia de tórax é capaz de fornecer grande quantidade de informações anatômicas e fisiológicas, e é considerada o padrão ouro para diagnóstico de pneumonia aguda [5]. Apesar de ser um exame simples, de ampla disponibilidade, a interpretação dos resultados muitas vezes é difícil, por diversos fatores, tais como a capacidade de interpretação dos profissionais, a qualidade dos equipamentos utilizados, as variações de idade e fisiológicas diversas dos pacientes.

A Indústria 4.0 [6] é caracterizada pela integração de diferentes tecnologias como Inteligência Artificial (IA), robótica, computação em nuvem, etc. As Redes Neurais Convolucionais (CNNs) são métodos eficazes para extrair características de imagens [7]. A aprendizagem profunda é um ramo da IA e tem aplicações em diversas áreas, como na área da saúde para detecção de anomalias em imagens médicas, e pode auxiliar profissionais da saúde na tomada de decisões, permitindo um diagnóstico mais rápido e preciso [8].

A radiografia de tórax é uma ferramenta fundamental para o diagnóstico de doenças, pois permite que os médicos monitorem e avaliem as condições físicas do paciente. Algoritmos de aprendizado de máquina estão cada vez mais sendo utilizados na interpretação e detecção de doenças [9]. Isso permite que médicos e profissionais de saúde tenham previsões sobre o diagnóstico do paciente, com isso realizar triagens de doenças e tomar decisões sobre o tratamento, melhorando a sensibilidade e a especificidade dos resultados.

Atualmente pesquisas apontam que a transferência de aprendizado de redes neurais pré-treinadas são alternativas eficientes

para modelagem preditiva de alto desempenho [7], e entre as redes neurais amplamente utilizadas estão: ResNet-50, VGG-19, DenseNet201, Inception-V3, MobileNet, Xception, etc.

A necessidade de desenvolver métodos computacionais rápidos, eficientes e precisos para diagnosticar doenças como a pneumonia é de interesse público. O objetivo deste trabalho foi classificar imagens de radiografia de tórax para identificar pacientes com pneumonia. Para isso foi utilizada a rede neural ResNet-50 para a extração de características e o modelo de aprendizado de máquina Random Forest para classificação.

Na Seção II, são discutidos tópicos relevantes de pesquisa. Na Seção III, são apresentados os referenciais teóricos utilizados no artigo, abrangendo redes neurais, modelagem preditiva, técnicas de otimização de parâmetros e metodologias de avaliação de modelos. A Seção IV trata do conjunto de dados e da origem das imagens empregadas. Os resultados do modelo nos processos de validação cruzada, treino, teste e validação estão expostos na Seção V. Na Seção VI, são delineadas as conclusões do estudo, bem como as possibilidades para pesquisas futuras.

II. REVISÃO DA LITERATURA

O desenvolvimento de sistemas baseado em IA capazes de identificar padrões e diagnosticar pacientes com diferentes patologias utilizando imagens é um tema presente na literatura. Em [10] os autores desenvolveram uma metodologia para diagnosticar pacientes com pneumopatia pediátrica com auxílio de imagens de radiografia de tórax. Esse sistema também é capaz de fazer a triagem de pacientes com doenças tratáveis comuns da retina que causam cegueira. O conjunto de dados consiste em imagens de Tomografia de Coerência Óptica e de tórax de pacientes. Neste trabalho eles utilizam uma abordagem de aprendizado por transferência da CNN Inception-V3. O desempenho da rede neural desenvolvida é comparável à classificação de especialistas humanos para diagnóstico dessas doenças de degeneração macular relacionada à idade e edema macular diabética. No trabalho [11] os autores desenvolveram uma CNN baseada na arquitetura da VGG-16 para classificação de imagens de Tomografia de Coerência Óptica, a técnica demonstrou a capacidade de distinguir imagens de degeneração macular relacionada à idade comparando com a de indivíduos saudáveis.

Em [6] os autores utilizaram o modelo Random Forest para detecção de anomalias e previsão do peso do pneu de uma fábrica. Nesse estudo um algoritmo de evolução diferencial foi utilizado para otimizar parâmetros de modelos de aprendizado de máquina e isso possibilitou melhores resultados comparado a outros métodos. O modelo Random Forest foi utilizado para classificar imagens com base em características extraídas da CNN ResNet-50 [12]. Esse estudo utilizou dados de imagens de retinopatia diabética e os autores relatam que a proposta adotada superam arquiteturas como: ResNet-50, VGG-19, Inception-v3, MobileNet, Xception e VGG-16.

No artigo [9] é apresentado um método não supervisionado para agrupar imagens de radiografia de tórax com base em medições de similaridade e dissimilaridade. As informações

das imagens foram extraídas usando uma CNN, e essas variáveis foram utilizadas para projeção dos dados com algoritmo Self-Organizing Map (SOM) em função de variáveis provenientes do processo de redução de dimensionalidade com a análise de componentes principais (PCA). Os autores adotaram o algoritmo K-means para criar os grupos. Isso permitiu o desenvolvimento de uma plataforma de triagem para detecção de tuberculose.

A ResNet-18 foi utilizada para extrair características de imagens de radiografia de tórax de pacientes com tuberculose pulmonar [13]. Os autores desenvolveram uma metodologia baseada em máquinas de comitê (*ensemble*) para classificação de imagens com diferentes configurações e padrões de qualidade. Nesse estudo os autores adotaram a técnica PCA para projeção dos dados em dimensão menor. A rede neural *multilayer perceptron* (MLP) foi adotada como classificador e o treinamento do modelo foi realizado com processo de validação cruzada usando k-folds.

O aumento de dados sintéticos utilizando técnicas de *data-augmentation* com auxílio de redes adversárias generativas (GANs) foi tema dos autores [14]. Nesse trabalho foi realizado uma comparação entre o desempenho de uma CNN com modelo de aprendizado de máquina *One Class Support Vector Machine* para classificação de imagens de pacientes com tuberculose. Os autores relatam ganhos no desempenho do treinamento quando os dados sintéticos foram gerados pela GAN *Pix2Pix*.

Os trabalhos [15], [16], [17], [18] apresentam diversas arquiteturas de redes neurais pré-treinadas para classificação de radiografia de tórax de indivíduos com COVID-19. A transferência de aprendizado com a rede ResNet-50 foi a arquitetura adotada para detecção de COVID-19 em [19]. Nesse artigo foram testadas várias composições de pesos diferentes da rede. Os resultados dos melhores modelos são da ordem de 99% de *accuracy*, *precision* e *recall*. No trabalho [20] a ResNet-50 teve o menor desempenho para classificação radiografia de tórax de pacientes com COVID-19. Os autores relatam que a arquitetura da VGG-16 forneceu o melhor desempenho com 80% de *accuracy*. Esse resultado foi comparado com as redes VGG-19, ResNet-50, DenseNet e Inception-V3.

Os autores do trabalho [21] também apresentaram uma metodologia baseada na técnica GAN para melhorar a triagem de pacientes com COVID-19. O conjunto de dados consiste em imagens de radiografia de tórax. Os autores discutem vários métodos de filtros para pré-processamento dos dados, como: Sobel, Laplaciano do Gaussiano e Gabor. Os resultados do treinamento da rede com a estratégia dos filtros foram comparados com as arquiteturas: AlexNet, GoogleNet, VGG-19, ShuffleNet-V2, DenseNet-121 e DenseNet-201. A metodologia com filtro de Gabor atingiu o melhor desempenho, com apenas 45 épocas o modelo foi capaz de elevar a precisão em até 32%.

Para a classificação de imagens de ressonância magnética de pacientes com COVID-19 com tecidos infectados, os autores [22] propuseram uma metodologia para o pré-processamento das imagens baseada em fractais, que consiste em reduzir a

dimensão e selecionar características relevantes das imagens. Essas informações serviram de entrada para o treinamento de uma rede neural profunda (DNN) e uma CNN proposta pelos autores. A classificação dos resultados mostra que a arquitetura da CNN apresentou os melhores resultados, com *accuracy* de 93,2%, comparado com *accuracy* de 83,4% do método DNN.

III. METODOLOGIA

Nesta sessão serão discutidos os referenciais teóricos das técnicas utilizadas para o desenvolvimento deste artigo. A Figura 1 representa a abordagem sequencial do tratamento dos dados. Os métodos foram implementados com auxílio de uma máquina com processador Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz 2.90 GHz, intel core i7 7th Gen, 8 GB de RAM e uma placa de vídeo NVIDIA GEFORCE 2 GB. A linguagem de programação adotada foi Python versão 3.9, com auxílio de ferramentas de software livre como Jupyter Notebook. As bibliotecas TensorFlow, scikit-learn e sklearn-genetic-op foram utilizadas para aplicação da ResNet-50, modelagem com Random Forest e otimização de parâmetros, respectivamente.

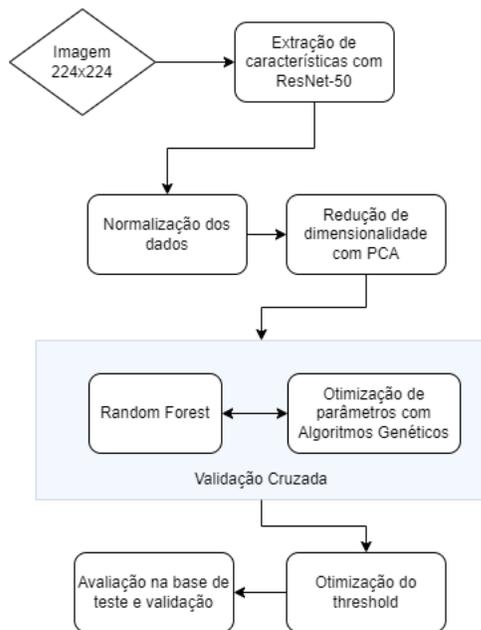


Figura 1. Fluxograma da metodologia.

A. Extração de características

A extração de características é fundamental no processo de classificação de imagens, e esse processo pode ser realizado por meio de uma CNN [19]. O treinamento de redes neurais profundas é um desafio para comunidade científica, e conforme o número de camadas de um modelo aumenta, o número de parâmetros no modelo aumenta significativamente, isso gera uma complexidade e requer recursos computacionais eficientes associados a processamento e memória [23]. Um dos problemas comuns em CNNs é a explosão e o desaparecimento dos gradientes. No processo de retropropagação, a diferenciação da regra da cadeia pode gerar valores de gradientes próximos

de zero, e quando multiplicados por mesma magnitude esse valor fica menor, analogamente o mesmo acontece para valores grandes, quando multiplicados os pesos podem gerar valores na ordem dos milhões, consequentemente gera instabilidade numérica dos parâmetros da rede.

A ResNet-50 é uma rede treinada com milhares de imagens do conjunto de dados ImageNet e contém 50 camadas [19]. As características extraídas das camadas mais profundas codificam propriedades específicas, como forma, textura e cor. A ResNet-50 trabalha com conceito de rede residual, permitindo conexões de saltos de alguma camada para outra camada oculta de modo a incorporar o fluxo de informações [12]. O processo permite que as camadas se encaixem em um mapeamento residual, conforme a Figura 2. O diferencial da ResNet comparada a outras redes é que esse mapeamento permite um menor custo computacional, além de demonstrar que o resultado da rede não necessariamente está associado a um número grande de camadas ocultas [24]. Os neurônios das camadas ocultas são transformados usando alguma função de ativação, as funções comumente utilizadas no contexto de classificação são: sigmoid; ReLu; tangente hiperbólica; softmax [15]. A equação 1 corresponde a forma como a derivada da função ReLu é implementada, onde x representa o valor do dado.

$$ReLU'(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (1)$$

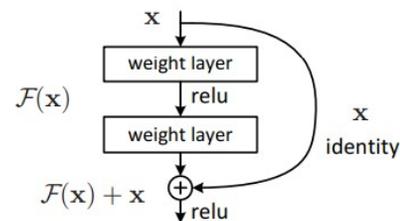


Figura 2. Rede residual da ResNet [24].

A Figura 3 representa a arquitetura da rede, ela contém uma série de camadas convolucionais (conv), as imagens de entrada foram padronizadas para dimensão (224, 224), as camadas intermediárias são processadas de 3 a 6 vezes. A camada de *average pool* é responsável por reduzir os dados de entrada, para isso é construído um mapa de recursos com a finalidade de agrupar as informações das camadas anteriores. No presente trabalho para extração das características são consideradas as transformações até o *average pool* da ResNet-50.

B. Análise de Componentes Principais

A análise de componentes principais (PCA) é uma técnica estatística que pode ser utilizada para explicar relações entre variáveis [25]. O método consiste em transformar os dados originais em outro conjunto de dimensão igual ou menor. O conjunto de componentes principais é formado por combinações lineares das variáveis originais [26]. No caso bidimensional,

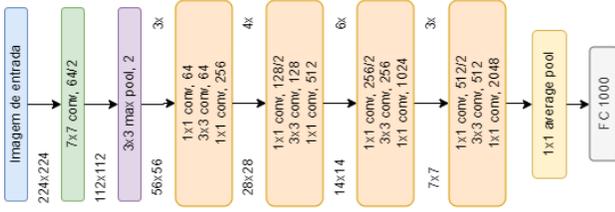


Figura 3. Arquitetura da ResNet-50

a primeira componente principal pode ser expressa conforme a equação 2 de modo a maximizar a variância expressa pela equação 3. O vetor \mathbf{X} corresponde às variáveis explicativas, a matriz de covariância é dada por Σ , e β são os coeficientes. A segunda componente principal é obtida por a equação 4, de modo a maximizar $\beta_2^T \Sigma \beta_2$, com as restrições $\beta_2^T \beta_2 = 1$ e $\beta_1^T \beta_2 = 0$. Portanto, a soma das variâncias das componentes principais é igual a variância total do sistema conforme a equação 5, e as variâncias denotadas por $\hat{\lambda}_i$ correspondem autovalores da matriz \mathbf{K} , e os coeficientes $\hat{\beta}_i$ são os autovetores [27].

$$PC_1 = \beta_1^T \mathbf{X} = \beta_{1,1} X_1 + \dots + \beta_{1,n} X_n. \quad (2)$$

$$V(PC_1) = \beta_1^T \Sigma \beta_1. \quad (3)$$

$$PC_2 = \beta_2^T \mathbf{X}. \quad (4)$$

$$tr(\mathbf{K}) = tr\left(\sum_{i=1}^2 \hat{\lambda}_i \hat{\beta}_i \hat{\beta}_i^T\right). \quad (5)$$

C. Random Forest

Os métodos de árvores de decisão são eficazes para análise preditiva e são mais fáceis de interpretar. Árvores de decisão são modelos de aprendizado de máquina supervisionado que representam regras de decisão baseadas nos valores dos atributos. O algoritmo de *Random Forest* (RF) proposto por [28] é composto por uma combinação de árvores de decisão, que cada árvore é sintetizada a partir de um vetor aleatório amostrado de forma independente com mesma distribuição de probabilidade.

A construção de uma árvore de classificação consiste em determinar as regiões em que o espaço das variáveis preditoras é particionado, com a finalidade de prever ou classificar uma variável resposta a partir de um conjunto de variáveis preditoras. As árvores do ensemble são construídas por processo de amostragem bootstrap. As florestas aleatórias adotam o bagging proposto com objetivo de reduzir a variância através da predição de vários modelos separados usando diferentes conjuntos dos dados de treinamento. É adotada amostragem com reposição com dois terços dos dados, para depois calcular a média dos resultados das predições. Os dados out-of-bag (OOB) são usados para validação e para definição de relevância dos atributos de entrada, e correspondem a um terço dos dados. As métricas de Gini e de entropia, conforme as

equações 6 e 7, são utilizadas como critério para escolha da variável que compõe o nó raiz de cada árvore, onde \hat{p}_k é a probabilidade da classe k . A entropia é uma medida de informação que indica a desordem do sistema, o Gini é uma medida de impureza, ambos possuem valores de mínimo igual a zero, isso indica que o nó da árvore é puro.

$$Entropia = \sum_{i=1}^k -\hat{p}_k \log(\hat{p}_k). \quad (6)$$

$$Gini = \sum_{i=1}^k \hat{p}_k (1 - \hat{p}_k). \quad (7)$$

Para problemas de classificação, o erro de generalização depende da ponderação atribuída às árvores individuais e da correlação de saída que há entre essas árvores. No processo de votação das árvores a classe majoritária vence entre as classes previstas [25]. Trata-se de um algoritmo robusto e eficiente para lidar com *overfitting*, o que o torna um modelo excelente para lidar com padrões não lineares como características extraídas da ResNet-50 [12].

D. Algoritmo Genético

O Algoritmo Genético (AG) é um tipo de algoritmo inspirado na evolução natural que é usado em problemas de otimização e busca em geral [29]. Os cromossomos representam essas soluções e podem ser codificados por uma matriz de bits. O conceito de geração, função *fitness*, cruzamento, mutação e elitismo são elementos fundamentais do AG. A geração é caracterizada por uma interação que contém uma população de cromossomos. A seleção dos cromossomos com melhores características está associado a função *fitness* [30]. O cruzamento consiste em selecionar dois cromossomos e gerar um cromossomo com uma nova configuração. O conceito de mutação serve para diversificar os indivíduos da população de modo que os cromossomos gerados no processo de cruzamento não fiquem presos em mínimos locais, esse operador altera com uma probabilidade baixa um dos genes do cromossomo aleatoriamente [31]. O elitismo consiste em selecionar os melhores indivíduos de cada geração, a seleção dos cromossomos pais para o processo de cruzamento está associado a uma distribuição de probabilidade com indivíduos com maior *fitness* da geração. O método GAsSearchCV [32] da biblioteca *sklearn-genetic-op* foi utilizado para estimar parâmetros do modelo *Random Forest* para maximizar a função *fitness*. Esse método seleciona conjuntos aleatórios de parâmetros e ajusta um modelo para cada conjunto de hiperparâmetros, para cada processo da validação cruzada é calculado o valor da função *fitness*. O algoritmo cria novas gerações combinando a última geração com diferentes abordagens, o processo é repetido até que há convergência considerando o número de gerações ou algum método de parada. A Figura 4 representa o fluxograma desse processo.

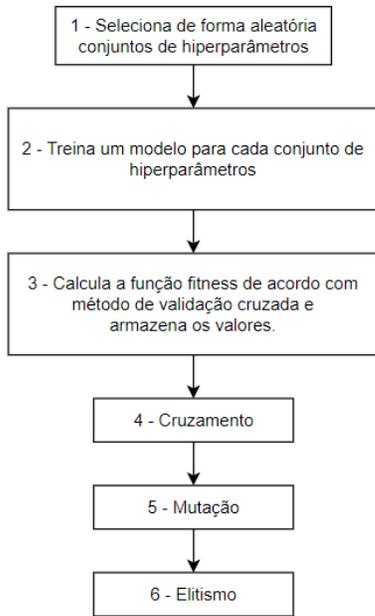


Figura 4. Fluxograma do GASearchCV.

E. Métricas de Desempenho

A avaliação do desempenho do modelo ajustado será baseado nas métricas derivadas da matriz de confusão M [33]. *Accuracy* conforme a equação 8 é correspondente a proporção de predições corretas do modelo. *Precision* expressa pela equação 9 é definida como o número de verdadeiros positivos (*True Positives* - TP) sobre o número de verdadeiros positivos somado ao número de falsos positivos (*False Positives* - FP). *Recall* conforme a equação 10 é definido como o número de verdadeiros positivos sobre o número de verdadeiros positivos somado o número de falsos negativos (*False Negatives* - FN). *Specificity* expressa pela equação 11 corresponde ao número de verdadeiros negativos (*True Negatives* - TN) sobre número de verdadeiros negativos somado ao número de falsos positivos. O F_1 -score conforme a equação 12 é a média harmônica do *precision* e *recall*. A área sob a curva característica de operação do receptor ROC-AUC foi utilizada como critério de avaliação, em que o eixo das abcissas contém a informação da taxa de falsos positivos, e o eixo das ordenadas a taxa de verdadeiros positivos.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$Specificity = \frac{TN}{TN + FP} \quad (11)$$

$$F_1score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (12)$$

A técnica k-fold foi utilizada no estudo para validação cruzada. Essa metodologia divide o conjunto de dados em k partes, em cada ajuste uma dessas partes é considerada como conjunto de testes e as demais como treinamento [25].

IV. ESTUDO DE CASO

O conjunto de dados utilizado está disponível em [34], e este possui 5863 imagens de Raios X com três categorias, normal, pneumonia bacteriana e pneumonia viral. As imagens de radiografia de tórax foram selecionadas de cortes retrospectivos de pacientes pediátricos do *Guangzhou Women and Children's Medical Center*. Os diagnósticos das imagens foram classificados por dois médicos especialistas. Para desenvolvimento desse estudo, foi feita a separação do conjunto em duas categorias, normal e pneumonia. O conjunto de amostras de 5216 imagens foi separado para treinamento do modelo, com 74,3% classificadas com pneumonia e 25,7% normal. O conjunto de teste contém 624 imagens com 62,5% e 37,5%, respectivamente, pneumonia e tórax normal. O conjunto de validação contém com 16 imagens, com 8 de cada categoria. Na Figura 5 temos um exemplo de imagens do conjunto de dados.



(a) Pulmão normal (b) Pulmão com pneumonia

Figura 5. Imagem 5(a) corresponde a uma radiografia de tórax normal e mostra pulmões claros, a imagem 5(b) exibe um pulmão com pneumonia.

V. RESULTADOS E DISCUSSÃO

A padronização nos dados foi aplicada devido a diferença entre os intervalos das variáveis preditoras, conforme a equação 13. O método consiste em escalar as variáveis de modo que a média seja zero e desvio padrão igual a 1.

$$Z = \frac{x - \mu}{\sigma} \quad (13)$$

O PCA permitiu que o estudo fosse realizado com um número de dimensão menor, com 600 componentes foi possível explicar 91,5% da variância total explicada pelas 2048 características extraídas da ResNet-50. Isso permitiu um tempo de processamento menor para o treinamento do modelo, sem perda de informações relevantes.

Na otimização de parâmetros do RF com AG foi considerado: uma população de tamanho 8; 5 gerações; probabilidade

de 0,8 de crossover; probabilidade de 0,1 de mutação; função fitness definida pela ROC-AUC. O método de validação cruzada k-fold consiste em amostras aleatórias estratificadas, ou seja, preservando a porcentagem de amostras para cada classe, com $k = 5$.

O intervalo de busca dos parâmetros do RF foram: número de árvores na floresta ficou entre $[20, 80]$, a profundidade máxima da árvore em $[4, 8]$, para medir a qualidade de uma divisão os critérios gini e entropy foram considerados, as funções 'sqrt' e 'log2' foram utilizadas para determinar a quantidade de variáveis predictoras na amostra, os intervalos para o número mínimo de amostras necessárias para divisão de um nó interno e para estar em um nó folha foram, respectivamente, $[80, 150]$ e $[200, 350]$.

A Figura 6 representa a quantidade de modelos ajustados e para cada interação o valor da média da métrica ROC-AUC nos conjuntos de treinamentos e validações no processo de validação cruzada. A margem sobre as linhas do gráfico corresponde ao desvio padrão de cada modelo em função dos k-fold's. A Figura 7, corresponde ao ROC-AUC em função do número de gerações do processo de otimização de parâmetros do AG, é possível notar que a partir da terceira geração o desempenho do modelo não tem diferença significativa, com ROC-AUC de 97,8% na base de treinamento. Os parâmetros do melhor modelo no processo de validação cruzada foram: a entropia para o critério de divisão; profundidade da árvore igual a 6; número mínimo de amostras de 88 para nó folha; 218 para divisão de um nó interno; e 75 para número de árvores.

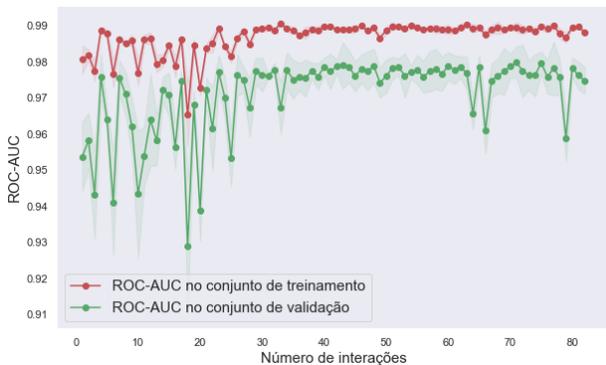


Figura 6. ROC-AUC em função do número de interações do processo de validação cruzada.

Para avaliar a capacidade de generalização do modelo a métrica de desempenho foi analisada em função do número de amostras para treinamento, para isso foi selecionado o melhor modelo do resultado de otimização com AG e foi avaliado em outro processo de validação cruzada. Na Figura 8 representa as curvas de aprendizado do RF e funciona como um mecanismo para diagnosticar o problema de viés e variância do modelo. Nota-se que conforme o número de amostras aumenta os valores de ROC-AUC do conjunto de validação cruzada convergem para os valores de inferência no conjunto de treinamento. Podemos observar que não há

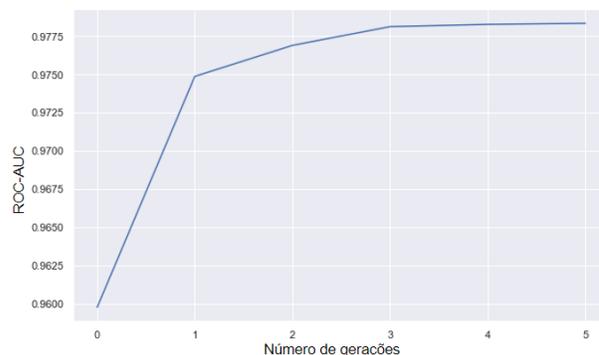


Figura 7. ROC-AUC em função do número de gerações.

muita diferença entre as curvas, o que corrobora na tese de que o modelo consegue generalizar sua inferência para outros conjuntos de dados.

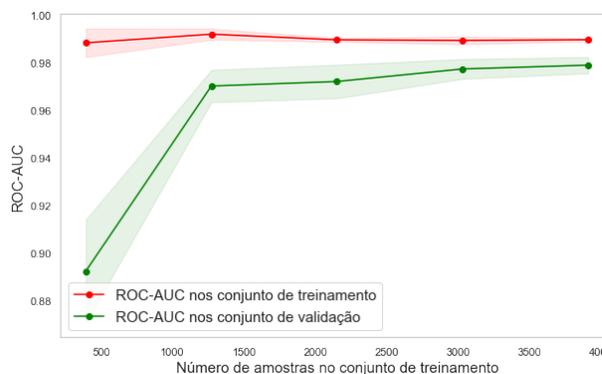


Figura 8. ROC-AUC do conjunto de treino e validação em função do número de amostras.

Para inferência nos conjuntos de dados de teste e validação foi optado por otimizar o ponto de corte do modelo em função da distribuição de probabilidade. Conforme a Figura 9 pode-se observar que a distribuição de probabilidade é assimétrica, a classe de radiografia de tórax com pneumonia é concentrada próximo ao intervalo de $[0,8; 1,0]$, a classe de imagens normais está distribuída no intervalo $[0,5; 0,6]$. A fim de obter melhores métricas, escolhemos o ponto de corte de 0,7 para probabilidade que distingue ambas as classes conforme a frequência observada no conjunto de treinamento.

As matrizes de confusão estão representadas na Figura 10, as métricas obtidas podem ser verificadas na Tabela I. O classificador RF obteve o desempenho de ROC-AUC de 93%, 79% e 88% nos conjuntos de treino, teste e validação, respectivamente. O tempo de processamento para o treinamento do modelo foi de 12,3 minutos. O modelo obteve os menores desempenhos da métrica *specificity* nos conjuntos de teste e validação, esse resultado indica que a proporção de falsos positivos aumentou em relação a proporção observada no conjunto de treinamento, o que não é um problema pois é preferível que o modelo tenha alto *recall* de modo a encontrar a maior parte dos pacientes doentes, mesmo que classifique

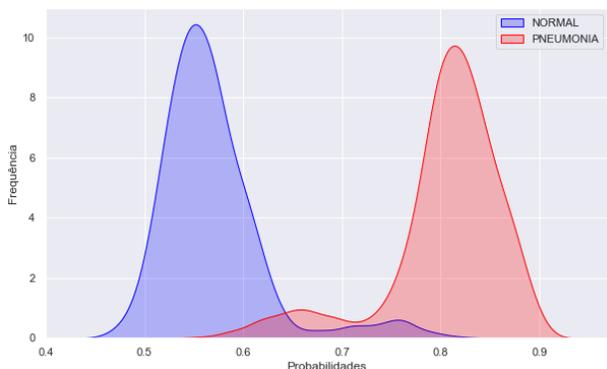


Figura 9. Distribuição da probabilidade do modelo Random Forest.

alguns pacientes saudáveis como doentes.

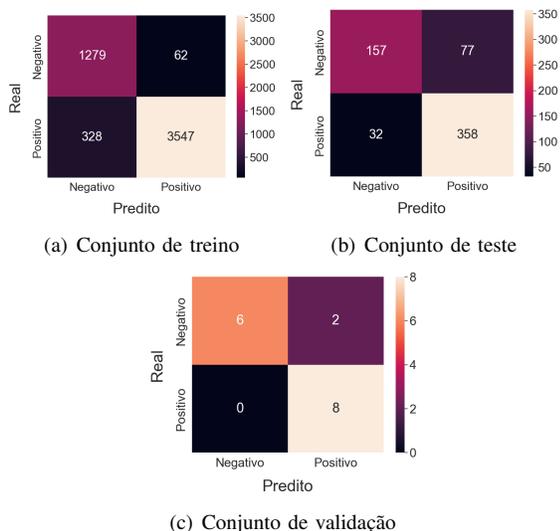


Figura 10. Matriz de confusão para cada conjunto de dados

 Tabela I
MÉTRICAS DE DESEMPENHO

Métricas de avaliação	Conjunto de dados		
	Treino	Teste	Validação
Accuracy	93%	83%	88%
Precision	98%	82%	80%
Recall	92%	92%	100%
Specificity	95%	67%	75%
F_1 -score	95%	87%	89%
ROC-AUC	93%	79%	88%

Comparando os resultados obtidos por outros trabalhos que abordam a classificação de imagens de radiografia de tórax, nota-se que o desempenho alcançado pelo modelo proposto é comparável com resultados vistos na literatura. Em [10] os autores utilizaram o mesmo conjunto de dados, e alcançaram um valor de *accuracy* de 93%, com *specificity* de 90% e *recall* de 93% no conjunto de teste. Os autores [15] relatam que o tempo de treinamento das redes pré-treinadas como ResNet-

50, Inception-V3 e ResNet-101 são em média da ordem de 4 horas. É evidente que a transferência de aprendizado dessas redes pré-treinadas contribuem significativamente no desempenho da classificação das imagens. Os resultados vistos em [15], [16], [17], [18] são da ordem de 99% de *accuracy*. Contudo, o tempo de execução para treinamento é relativamente maior comparado a abordagem adotada nesse trabalho. A estratégia de usar PCA para reduzir a dimensão das variáveis predictoras se demonstra factível para inferência do modelo. O resultado obtido nesse trabalho demonstra que a metodologia é robusta e comparável aos resultados da literatura, embora os valores de algumas métricas de desempenho sejam inferiores comparado a alguns trabalhos elaborados para classificação de imagens de radiografia de tórax, o modelo consegue distinguir com alto desempenho imagens de indivíduos saudáveis e com pneumonia. A proposta tem um tempo de treinamento relativamente baixo, o que pode ser interessante para testar novas arquiteturas de redes neurais e modelos de aprendizado de máquina que confronte os resultados apresentados pela literatura.

VI. CONCLUSÃO

Os modelos de aprendizado de máquinas são técnicas amplamente utilizadas para classificação e detecção de padrões não lineares. O presente trabalho apresentou algoritmos de aprendizado de máquina utilizado para classificação de imagens de radiografia de tórax de indivíduos saudáveis e com pneumonia. A rede neural pré-treinada ResNest-50 foi utilizada para extrair características das imagens. Esse processo resultou em 2048 variáveis predictoras. Com auxílio da PCA foi possível reduzir a dimensão para 600 componentes principais, com 91% da variância total. Essas informações foram utilizadas para classificação das imagens usando o modelo RF em conjunto com AG no processo de validação cruzada para otimização dos parâmetros. Foi realizado uma análise de otimização do melhor limiar de probabilidade para classificação das imagens.

Os resultados desse trabalho contribuem para desenvolvimento de sistemas de IA com alto desempenho. O processamento dessas informações pode contribuir para agilizar o diagnóstico de doenças tratáveis como a pneumonia, facilitando o tratamento precoce. O modelo proposto obteve um desempenho de F_1 -score de 95%, 87% e 89% nos conjuntos de treino, teste e validação, respectivamente, com tempo de execução baixo. O treinamento do modelo foi realizado com uma máquina comum, o que torna o experimento alcançado seja facilmente reproduzível. Essa foi a razão para não usar uma CNN para classificação, pois o treinamento da rede neural requer recursos computacionais para processamento de alto desempenho, conforme os resultados vistos na literatura as tecnologias como *Graphics Processing Unit* (GPU) são fundamentais para reduzir o tempo de treinamento das CNNs [15]. Para pesquisas futuras se pretende avaliar outros algoritmos de aprendizado de máquina, em conjunto com outras arquitetura de redes neurais pré-treinadas, como VGG-19, DesNet-101, CoroNet, ResNest-101, etc [21] [20]. A avaliação

do desempenho do modelo entre os grupos com pneumonia viral e bacteriana pode contribuir para demonstrar a robustez dos métodos abordados, embasando o desenvolvimento de modelos para classificar o tipo de pneumonia em indivíduos doentes.

REFERÊNCIAS

- [1] R. k. Seramo, S. M. Awol, Y. A. Wabe, and M. M. Ali, "Determinants of pneumonia among children attending public health facilities in Worabe town," *Scientific Reports* 2022 12:1, vol. 12, pp. 1–9, 4 2022.
- [2] J. C. Rodrigues, L. V. F. Da Silva Filho, and A. Bush, "Diagnóstico etiológico das pneumonias: uma visão crítica," *Jornal de Pediatria*, vol. 78, pp. 129–140, 11 2002.
- [3] R. A. M. Menezes, D. R. Pavanitto, and L. F. C. Nascimento, "Distribuição espacial das taxas de internação de crianças por pneumonia no Sistema Único de Saúde, nos municípios do estado de São Paulo," *Revista Brasileira de Epidemiologia*, vol. 22, p. e190053, 12 2019.
- [4] R. d. O. Ferraz, J. K. Oliveira-Friestino, and P. M. S. B. Francisco, "Pneumonia mortality trends in all Brazilian geographical regions between 1996 and 2012," *Jornal Brasileiro de Pneumologia*, vol. 43, pp. 274–279, 7 2017.
- [5] S. Raoof, D. Feigin, A. Sung, S. Raoof, L. Irugupati, and E. C. Rosenow, "Interpretation of plain chest roentgenogram," *Chest*, vol. 141, pp. 545–558, 2 2012.
- [6] S. M. Acosta, R. M. A. Oliveira, and M. O. Sant'Anna, "Machine learning algorithms applied to intelligent tyre manufacturing," *International Journal of Computer Integrated Manufacturing*, pp. 1–11, 2 2023.
- [7] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A. Alamri, "Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review," *Remote Sensing*, vol. 12, 5 2020.
- [8] G. Silva, L. Duarte, M. Shirassu, S. Peres, M. de Moraes, and A. Chiavegatto Filho, "Machine learning for longitudinal mortality risk prediction in patients with malignant neoplasm in São Paulo, Brazil," *Artificial Intelligence in the Life Sciences*, vol. 3, p. 100061, 12 2023.
- [9] F. Ferreira, P. Gaspar, L. M. Oliveira, R. Torres, M. V. Araújo, C. E. Covas, M. Bastos, A. Trajman, and J. M. Seixas, "Machine Learning based sampling of X-Ray images for a computer-aided detection of Tuberculosis," *Associacao Brasileira de Inteligencia Computacional*, pp. 1–7, 10 2021.
- [10] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M. K. Prasadha, J. Pei, M. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V. A. Huu, C. Wen, E. D. Zhang, C. L. Zhang, O. Li, X. Wang, M. A. Singer, X. Sun, J. Xu, A. Tafreshi, M. A. Lewis, H. Xia, and K. Zhang, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, pp. 1122–1131, 2 2018.
- [11] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images," *Ophthalmology Retina*, vol. 1, pp. 322–327, 2017.
- [12] M. K. Yaqoob, S. F. Ali, M. Bilal, M. S. Hanif, and U. M. Al-Saggaf, "ResNet Based Deep Features and Random Forest Classifier for Diabetic Retinopathy Detection," *Sensors (Basel, Switzerland)*, vol. 21, pp. 1–6, 6 2021.
- [13] C. A. S. Silva, L. A. A. Cardoso, D. D. Ferreira, J. M. Seixas, M. L. S. Bastos, and A. Trajman, "Ensemble de especialistas para avaliação de adesão ao procedimento operacional padrão de fotografias de radiografias de tórax," *Associacao Brasileira de Inteligencia Computacional*, pp. 1–7, 10 2021.
- [14] O. T. Nascimento, J. M. Seixas, and A. Trajman, "Data-augmentation de dados de radiografia de torax no contexto de aprendizagem profunda," *Associacao Brasileira de Inteligencia Computacional*, pp. 1–7, 10 2021.
- [15] A. Narin, C. Kaya, and Z. Pamuk, "Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks," *Pattern Analysis and Applications*, vol. 24, pp. 1207–1220, 3 2020.
- [16] A. I. Khan, J. L. Shah, and M. M. Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," *Computer methods and programs in biomedicine*, vol. 196, 11 2020.
- [17] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al-Emadi, M. B. I. Reaz, and T. I. Islam, "Can AI help in screening Viral and COVID-19 pneumonia?," *IEEE Access*, vol. 8, pp. 132665–132676, 3 2020.
- [18] R. d. S. Farias, K. R. d. S. Oliveira, C. D. M. Regis, C. d. M. Costa, and J. M. Seixas, "Redes Neurais Convolucionais para Classificação da COVID-19 em Imagens de Raio-x de Tórax," *Anais do 15. Congresso Brasileiro de Inteligencia Computacional*, pp. 1–8, 1 2021.
- [19] M. B. Hossain, S. M. S. Iqbal, M. M. Islam, M. N. Akhtar, and I. H. Sarker, "Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images," *Informatics in Medicine Unlocked*, vol. 30, p. 100916, 1 2022.
- [20] K. Sahinbas and F. O. Catak, "Transfer learning-based convolutional neural network for COVID-19 detection with X-ray images," *Data Science for COVID-19*, p. 451, 1 2021.
- [21] A. H. Barshooi and A. Amirkhani, "A novel data augmentation based on Gabor filter and convolutional deep learning for improving the classification of COVID-19 chest X-Ray images," *Biomedical Signal Processing and Control*, vol. 72, p. 103326, 2 2022.
- [22] S. Hassantabar, M. Ahmadi, and A. Sharifi, "Diagnosis and detection of infected tissue of COVID-19 patients based on lung x-ray image using convolutional neural network approaches," *Chaos, Solitons and Fractals*, vol. 140, 2020.
- [23] S. Haykin, "Neural Networks and Learning Machines," *Pearson Prentice Hall New Jersey USA 936 pLinks*, vol. 3, p. 906, 2008.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, 12 2015.
- [25] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Series in Statistics, New York, NY: Springer New York, 2nd ed ed., 2009.
- [26] M. M. Silva, E. F. Simas Filho, P. C. Farias, M. C. Albuquerque, I. C. Silva, and C. T. Farias, "Intelligent embedded system for decision support in pulsed eddy current corrosion detection using Extreme Learning Machine," *Measurement*, vol. 185, p. 110069, 11 2021.
- [27] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning : with Applications in R*. New York, NY: Springer, 2nd ed ed., 2013.
- [28] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 10 2001.
- [29] J. L. Carvalho, P. C. Farias, and E. F. Simas Filho, "Global Localization of Unmanned Ground Vehicles Using Swarm Intelligence and Evolutionary Algorithms," *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 107, pp. 1–21, 3 2023.
- [30] A. L. Pereira, M. Kohler, and M. A. C. Pacheco, "Evolutionary Convolutional Neural Network: a case study," *Associacao Brasileira de Inteligencia Computacional*, pp. 1–6, 11 2021.
- [31] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimedia Tools and Applications*, vol. 80, pp. 8091–8126, 2 2021.
- [32] G. Ruan, U. Schmidhalter, F. Yuan, D. Cammarano, X. Liu, Y. Tian, Y. Zhu, W. Cao, and Q. Cao, "Exploring the transferability of wheat nitrogen status estimation with multisource data and Evolutionary Algorithm-Deep Learning (EA-DL) framework," *European Journal of Agronomy*, vol. 143, 2 2023.
- [33] A. K. Santra and C. J. Christy, "Genetic Algorithm and Confusion Matrix for Document Clustering," 2012.
- [34] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification," vol. 2, 2018.