

Automatic Classification of Breast Lesions Using Transfer Learning

R. da Silva, and A. de Carvalho

Abstract—Breast lesions are one of the most common types of lesions among women in Brazil and worldwide, accounting for about 28% of new cases each year. These lesions may have Benign or Malignant behaviors. In this work, a computational methodology for image classification was developed to differentiate malignant and benign lesions breast, aiming at low computational cost and good efficiency. In our approach, different Convolutional Neural Networks architectures and several classifiers were tested. Transfer Learning was employed to deal with the limitation of the small number of images in the database, reaching an accuracy of 81.73%, a sensitivity of 85.66%, a specificity of 78.40%, Kappa of 0.63 and ROC curve of 0.82. Finally, it is believed that the proposed methodology can integrate a CAD tool acting as patient screening or providing a second opinion to the specialist.

Index Terms—Breast Lesions, Digital Image Processing, Transfer Learning, Deep Learning.

I. INTRODUÇÃO

Atualmente, o câncer de mama é uma das principais ocorrências de câncer em mulheres com mais de 40 anos de idade em todo o mundo. Um diagnóstico precoce é muito importante, e um método comum de diagnóstico é o uso da mamografia, que é basicamente um raio-X da região da mama tirada em uma condição especial. A partir da imagem, um médico treinado o examina buscando microcalcificações e massas de várias formas. Se encontrados, esses artefatos na imagem podem ser um sinal da presença de um tumor benigno ou maligno [1].

Ainda não é possível prevenir o câncer de mama em função da multiplicidade de fatores relacionados ao surgimento da doença, desta forma, o diagnóstico precoce é a melhor forma de combater a doença e fornecer um prognóstico mais favorável ao paciente. [2] aponta que o diagnóstico do câncer de mama, em países com recursos limitados, é frequentemente feito em um estágio já avançado.

Especificamente, durante o procedimento de diagnóstico, os especialistas avaliam a organização do tecido por meio de imagens microscópicas. No entanto, a grande quantidade de dados e a complexidade das imagens tornam essa tarefa custosa e não trivial. O método tradicional de inspeção microscópica é importantíssimo em analisar a estrutura do tecido. No entanto, esta técnica de inspeção manual é altamente demorada e propensa a erros manuais, principalmente no processo de classificação em lesões malignas ou benignas, pois, é nesse diagnóstico que é fornecido a informação de espalhamento de células cancerosas para os tecidos circundantes a partir

de seu ponto de ocorrência. Dessa forma, o desenvolvimento de ferramentas computacionais automáticas de diagnóstico é desafiador, mas também essencial para o campo da saúde.

Geralmente, os sistemas Computer Aided Diagnosis (CADx) são responsáveis por prover aos especialistas uma segunda opinião no diagnóstico médico, ou atuar na triagem de pacientes. Um CADx é composto por técnicas computacionais de processamento de imagens e inteligência computacional.

Pesquisas efetuadas mostraram que as técnicas de Deep Learning estão se tornando primordiais na resolução de diagnóstico médico por meio de imagens, por isso foi determinado o uso dessas técnicas envolvendo algumas arquiteturas de redes convolucionais. Um problema relacionado à técnicas de Deep Learning é seu alto custo computacional, além de ser preciso um longo tempo para ser treinada.

Uma abordagem que resolve tal problema é o uso do Transfer Learning [3], que tem como objetivo reduzir o custo computacional e o tempo de processamento. Tal técnica é eficiente em bases com pouca quantidade de imagens.

Nesse sentido, este artigo apresenta uma metodologia automática utilizando Transfer Learning com diversas arquiteturas de Convolutional Neural Network (CNNs) e múltiplos classificadores para classificação automática de lesões benignas e malignas da mama.

Em vista do contexto apresentado, este trabalho tem como objetivo geral desenvolver um método computacional de baixo custo computacional e eficiente para a classificação automática de imagens de lesões de mama em benignas e malignas em resolução ao problema de uma limitada base de imagens, reduzindo resultados falso positivo e falso negativo.

II. TRABALHOS RELACIONADOS

A Tabela I mostra um resumo das principais informações dos trabalhos relacionados para a classificação de lesões de mama, em seguida, apresentamos um resumo dos trabalhos relacionados.

TABELA I
COMPARAÇÃO ENTRE OS TRABALHOS RELACIONADOS

Trabalho	Base de Imagens	Algoritmos	Acurácia
[4]	ICIAR 2018	ResNet, VGG e Inception	87,2%
[5]	ICIAR 2018	GoogleNet	80,0%
[6]	DDSM	LBP	90,18%
[7]	BCDR-FM	CNN, HOG e HGD	–

Em [4], foi proposto uma abordagem computacional baseada em redes neurais profundas, onde são utilizadas algumas arquiteturas de redes (ResNet, VGG e Inception) pré-treinadas

R. P. da Silva Neto, Universidade Federal do Piauí (UFPI), Picos, Piauí, Brasil, raahpsn@gmail.com.

A. O. de Carvalho Filho, Universidade Federal do Piauí (UFPI), Picos, Piauí, Brasil, antonioseas@ufpi.edu.br.

na base ImageNet para a classificação de imagens histológicas de câncer de mama da base International Conference on Image Analysis and Recognition (ICIA) 2018.

[5] propôs uma estratégia que utiliza CNN híbrida que é projetada com base na arquitetura GoogleNet para mesclar mais informações importantes na decisão, onde a Transferência de Aprendizado e aumento de dados foram empregados para lidar com as limitações da pequena quantidade de imagens na base ICIAR 2018.

Em [6], utilizando a base Digital Database for Screening Mammography (DDSM), foram combinadas abordagens estrutural e estatística para a análise de textura. A abordagem estrutural foi realizada por meio da técnica de LBP e as abordagens estatísticas, através das técnicas da função K de Ripley e de índice de diversidade.

O método proposto por [7], apresenta uma metodologia para classificação automática das lesões de imagem da mama utilizando representação do aprendizado. O método é avaliado de dados de benchmarking comprovado por biópsia 736 casos. O método desenvolvido compreende dois estágios principais: (i) pré-processamento para aprimorar os detalhes da imagem e (ii) treinamento supervisionado para aprender os recursos e o classificador de lesões de imagem da mama usando histogram of oriented gradients (HOG) e histogram of the gradient divergence (HGD). Os resultados apresentam desempenho de 0,787 a 0,822 em termos da área sob a curva ROC.

O principal diferencial desse trabalho é a análise dos principais modelos de CNNs disponíveis como descritores de características na base DDSM, utilizando múltiplos classificadores a fim de reduzir resultados falso positivo e falso negativo.

III. MATERIAIS E MÉTODOS

A primeira etapa foi a coleta das imagens de duas classes de lesões de mama, malignas e benignas. Para compor a base de dados utilizada nesse trabalho, foram selecionadas imagens do arquivo da DDSM, por ser de fácil acesso e aberta ao público.

A segunda etapa consistiu em extrair características dessas imagens utilizando as redes pré-treinadas com pesos da ImageNet, disponíveis na biblioteca Keras. Houve um processamento da base para alterar o tamanho de entrada dos dados para o formato exigido por cada uma das redes e, devido a transferência de aprendizado, as suas camadas de classificação foram removidas.

Na terceira etapa, foram utilizados os classificadores: Support Vector Machine (SVM), Random Forest e o eXtreme Gradient Boost (XGBoost), com as características extraídas das arquiteturas para a validação. Nesta etapa, foi utilizado o método cross validation. Foram testados ao todo três kernels diferentes do SVM: Linear, Polynomial e RBF, numa tentativa de alcançar maiores resultados.

Por fim, na quarta e última etapa, para a avaliação do método proposto nesse trabalho foram utilizadas as métricas: Sensibilidade (S), Especificidade (E), Acurácia (A), Área Sob a Curva ROC (AUC) e Kappa (K). Essas métricas fazem uso da matriz de confusão, que indica a classificação correta ou incorreta das classes em uso, agrupando os resultados em

função de quatro valores, sendo eles: Falso Negativo (FN), Falso Positivo (FP), Verdadeiro Positivo (VP) e Verdadeiro Negativo (VN).

A Figura 1 descreve um resumo das etapas do método utilizado.

A. Aquisição de Imagens

O Digital Database for Screening Mammography (DDSM) [8] é uma base de imagens públicas contendo 2.620 estudos/exames de mamografia digitais. Contém casos normais, benignos e malignos com informações patológicas confirmadas. A DDSM está disponível para uso da comunidade de pesquisa de análises mamográficas de imagens.

Todas os exames possuem marcações indicando os contornos das lesões. Estas marcações são realizadas por um conjunto de especialistas. Com isso, em nosso trabalho, utilizamos as marcações disponíveis para segmentar as Region of Interest (ROI) e com isso aplicar o método proposto.

B. Extração de Características

Tradicionalmente as mais variadas tecnologias são utilizadas para oferecer um ótimo desempenho no processamento de imagens e uma dessas está rapidamente se tornando um dos mais estudados e procurados campos dentro da ciência da computação moderna, onde ela oferece facilidades e poder computacional, tal tecnologia é denominado Deep Learning.

Deep Learning é um subcampo do aprendizado de máquina que tem como foco principal desenvolver meios para resolver problemas de forma semelhante ao processo de tomada de decisão humano.

Tradicionalmente, a qualidade dos algoritmos depende muito da representação dos dados em características. Assim, a criação recursos validos e pré-processamento consomem grande parte dos esforços dos especialistas. Já os algoritmos do tipo Deep Learning têm uma abordagem inovadora, pois dispensam grande parte desse pré-processamento e geram automaticamente propriedades invariantes em suas camadas hierárquicas de representação.

As CNN's consistem em camadas de convolução que são responsáveis pela extração de características, em camadas de *pooling* que subdivide as imagens em subgrupos de *pixels* cada vez menores e em camadas totalmente conectadas que funcionam como o classificador. Esse tipo de rede tem propriedade de compartilhamento dos pesos, o que possibilita o entendimento das operações entre pesos e dados de entrada como sendo operações de convolução. Sendo assim, considera-se que as CNNs possuem vários filtros que podem ser treinados para extrair características específicas dos dados [9].

Quanto maior a quantidade de amostras (base de dados) que representam as diferentes classes, mais fácil será o treinamento de uma CNN. Mas, nem sempre é possível utilizar grandes quantidades de imagens devido a empecilhos como barreiras de acesso a essas bases ou até mesmo limitação do hardware, o que atrapalha o uso dessa ferramenta. Para contornar esse problema, foi desenvolvido uma técnica chamada de Transfer Learning.

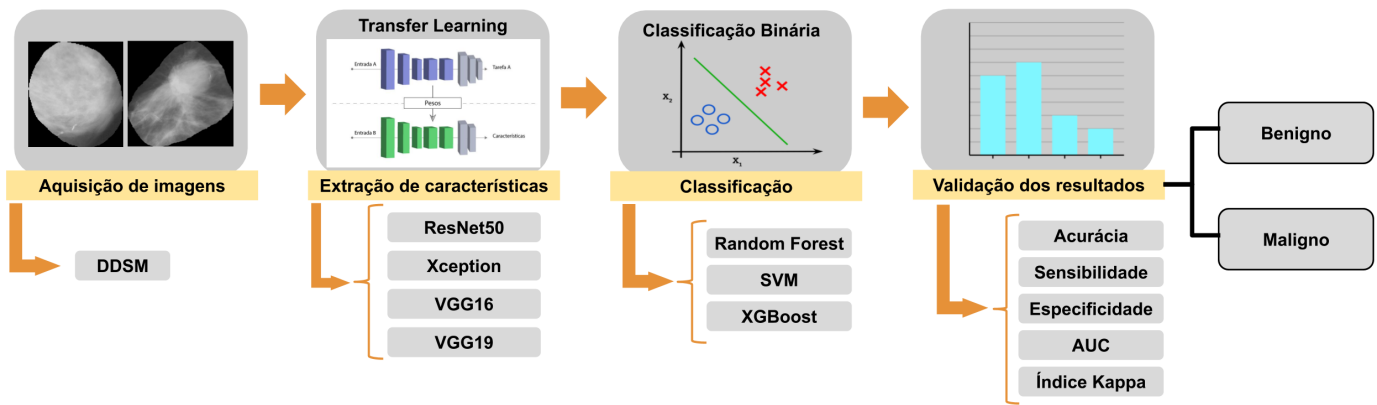


Fig. 1. Metodologia proposta.

C. Transfer Learning

O termo Transfer Learning se refere ao processo de transferir o conhecimento da tarefa de origem para uma tarefa de destino permitindo um progresso rápido ou duradouro ao modelar a segunda tarefa. Na área de aprendizagem de máquina essa técnica é utilizada quando não é possível obter uma base de dados grande o suficiente para resolver uma tarefa [10].

Normalmente, precisa-se de muitos dados para treinar uma CNN, mas nem sempre é possível ter acesso a dados suficientes. É aí que o Transfer Learning entra em jogo, porque com ela pode-se construir um modelo de aprendizado de máquina sólida com comparativamente poucos dados de treinamento, porque o modelo já é pré-treinado [11].

Embora o Deep Learning tenha mostrado-se promissor na representação de informações, tais modelos profundos supervisionados, exigem grandes quantidades de dados rotulados e recursos de computação, que são caros [12]. Foram selecionados três principais redes de CNN com pesos pré definidos para estudo e consequente avaliação da melhor performance as utilizando no conceito de Transfer Learning. Tais modelos são abordados a seguir.

1) **Modelo 1 - ResNet:** A ResNet é uma arquitetura que tem como principal característica suas conexões residuais. Na aprendizagem residual, em vez de tentar aprender algumas características, tal arquitetura é voltada para aprender alguns resíduos. Residual pode ser simplesmente entendido como subtração de características aprendidas da entrada dessa camada e a ResNet faz isso usando conexões de atalho, que são conexões em que o gradiente é diretamente retropropagado para camadas anteriores. Na Figura 2 é demonstrado o funcionamento de um bloco residual.

Um dos problemas que a ResNet resolve é o famoso desaparecimento de gradiente. Isso ocorre porque, quando a rede é muito profunda, os gradientes de onde a função de perda é calculada diminuem facilmente para zero após a aplicação da regra da cadeia por várias vezes. Este resultado nos pesos nunca atualiza seus valores e, portanto, nenhum aprendizado será realizado. Com tal arquitetura, os gradientes podem fluir diretamente através das conexões de salto. Para [13] a ResNet fornece uma estrutura de treinamento para facilitar formação de redes substancialmente mais profundas que as utilizadas

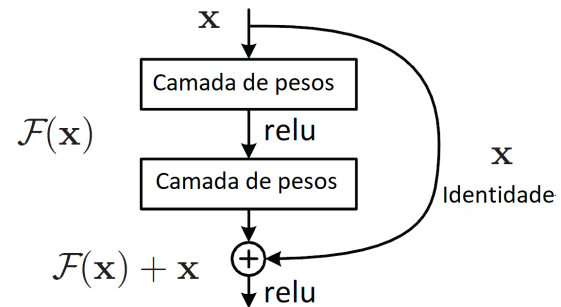


Fig. 2. Ilustração de um bloco residual na arquitetura da rede ResNet.

anteriormente.

2) **Modelo 2 - VGG:** A VGG é um modelo de CNN inicialmente proposto por [14]. A mesma garantiu a primeira e segunda posições nas tarefas de localização e classificação, respectivamente, no desafio ImageNet 2014. Ela faz a melhoria sobre o AlexNet construída por [15] substituindo grandes filtros do kernel por múltiplos filtros do tamanho 3x3 um após o outro. Nela as profundidades mais comuns são as de 16 e 19 camadas de profundidade.

Com a introdução da VGG, elas trouxeram alguns padrões: sugeriu-se que todos os filtros tivessem tamanho de 3x3, pools máximos deveriam ser colocados após cada 2 convoluções. A VGGNet é considerada a rede preferida pela comunidade para Transfer Learning, pois sua arquitetura uniforme é boa para extrair características de imagens. A Figura 3 nos mostra uma ilustração da arquitetura utilizada na VGG.

3) **Modelo 3 - Xception:** A arquitetura Xception foi proposta por [16] e significa *Extreme version of Inception* sendo uma versão da arquitetura Inception, onde essa arquitetura tem 36 camadas convolucionais formando a base de extração de características da rede. A sua principal característica é que ela substitui o módulo inception por convoluções separáveis em profundidade, isso faz com que o número de conexões sejam menores e o modelo se torne mais leve. A Figura 4 demonstra um módulo da rede Xception.

A Xception em vez de particionar dados de entrada em vários blocos compactados, a mesma leva isso um passo adiante, ela mapeia as correlações espaciais para cada canal de saída separadamente e, em seguida, executa uma convolução de 1x1

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Fig. 3. Ilustração da arquitetura da VGG.

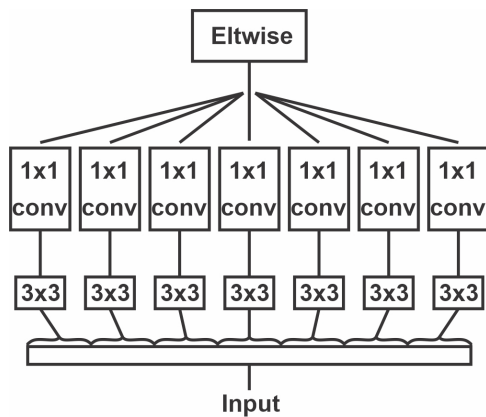


Fig. 4. Operações de convolução da Xception.

em profundidade para capturar a correlação entre canais.

D. Classificação

A classificação consiste em reconhecer novos objetos de um conjunto de recursos extraídos de um grupo de objetos de interesse e, em seguida, treinar um sistema para reconhecer novos objetos desses recursos [17]. Para esta classificação, foram utilizados: Support Vector Machine (SVM), Random Forest e eXtreme Gradient Boost (Xgboost).

- **SVM:** é um algoritmo de classificação de aprendizado de máquina supervisionado proposto por [18]. Neste trabalho, usamos o SVM com o kernel da função base radial (RBF), parâmetro $\gamma = 10$ e $C = 10000$. O valor de C foi escolhido alto para que uma pequena margem fosse aceita pela função de decisão na classificação correta de todos os pontos. O parâmetro γ foi escolhido baixo, de modo que um único exemplo teve pouca influência no treinamento de dados.

- **Random Forest:** é uma combinação de múltiplas previsões de árvores em que cada árvore depende dos valores de um vetor independente, sorteado aleatoriamente e com a mesma distribuição para todas as árvores da floresta [19]. Para este trabalho, o Random Forest foi usada com os seguintes parâmetros: bag Size Percent = 100, batch Size = 100, número de execuções slots = 1, max Depth = 0 (ilimitado), número de atributos escolhidos aleatoriamente = 0, número de iterações a ser realizado = 100, número mínimo de instâncias por folha = 1,0, variação mínima para divisão = 0,001 e o número aleatório semente a ser usado = 1.
- **Xgboost:** é uma biblioteca proposta por [20] com base na estrutura de aumento de gradiente desenvolvida por [21], [22]. O Xgboost pode ser usado para várias funções objetivas, como regressão e classificação [23]. O Xgboost é extensível e seus parâmetros podem ser facilmente alterados. Os parâmetros utilizados no XGBoost foram os seguintes: profundidade máxima = 7, taxa de aprendizado = 0,1, ite = 1000, $\gamma = 0$, passo delta máximo = 1 e objetivo = "multi:softmax".

E. Métricas de Validação

Para a avaliação do método proposto nesse trabalho foram utilizadas as seguintes métricas: Sensibilidade (S), Especificidade (E), Acurácia (A), Área Sob a Curva ROC (AUC) e $Kappa$ (K). Essas métricas fazem uso da matriz de confusão, que indica a classificação correta ou incorreta das classes em uso, agrupando os resultados em função de quatro valores, sendo eles: Falso Negativo (FN), Falso Positivo (FP), Verdadeiro Positivo (VP) e Verdadeiro Negativo (VN).

IV. RESULTADOS E DISCUSSÕES

Nesse trabalho, propomos um método automático para classificação de imagens de lesões de mama em duas classes (benigno e maligno) que foram coletadas da base DDSM usando abordagens de Transfer Learning, sendo 1155 ROIs, compostas de 530 lesões benignas e 625 malignas.

Foram utilizadas redes pré-treinadas com pesos da ImageNet [24] para extração das características dessas imagens. No processo final de classificação é utilizado os classificadores SVM com kernels (Linear, Radial basis function e Polynomial), Random Forest e o XGBoost, conforme descrito na Seção III-D.

O método de avaliação empregado nesse trabalho é o *Cross Validation*, onde consiste em dividir a base de imagens em k partes, usando $k-1$ partes para treino e a parte remanescente para teste, fazendo isso k vezes. Também foi feito a concatenação de todas as características extraídas pelos descritores para classificação, e por fim, fazemos um último teste que conta com a seleção de atributos utilizando o *Greedy StepWise* [25].

Para a realização dos experimentos aqui descritos, utilizou-se um computador, cujas especificações são descritas na Tabela II.

A seguir, os resultados obtidos com a metodologia proposta serão apresentados e discutidos. As métricas nas tabelas são: Acurácia (A), Sensibilidade (S), Especificidade (E), $Kappa$

TABELA II
DESCRIÇÃO DO AMBIENTE DE DESENVOLVIMENTO

Características	Descrição
Modelo	Acer Aspire VX5-591G-54PG
RAM	8,00 GB
Processador	Intel Core i5-7300HQ 2.5 GHz
Sistema Operacional	Windows 10

TABELA III
RESULTADOS DE TESTES NA REDE RESNET

Rede	Classificador	A(%)	S(%)	E(%)	K	AUC
Resnet50 Sem seleção 2048 features	RF	79,48	81,88	77,44	0,58	0,85
	SVM LINEAR	78,70	77,54	79,68	0,57	0,78
	SVM RBF	81,38	85,47	77,76	0,62	0,81
	SVM POLY	81,73	85,66	78,40	0,63	0,82
ResNet50 Com seleção 137 features	XGBoost	80,84	80,84	75,84	0,56	0,78
	RF	78,00	81,13	75,36	0,56	0,85
	SVM LINEAR	79,22	83,20	75,84	0,58	0,79
	SVM RBF	76,71	79,62	74,24	0,53	0,76
Com seleção 137 features	SVM POLY	78,96	85,47	73,44	0,58	0,79
	XGBoost	75,06	78,11	72,48	0,50	0,75

TABELA IV
RESULTADOS DE TESTES NA REDE XCEPTION

Rede	Classificador	A(%)	S(%)	E(%)	K	AUC
Xception Sem seleção 2048 features	RF	78,09	80,18	76,32	0,56	0,85
	SVM LINEAR	76,36	76,41	76,32	0,52	0,76
	SVM RBF	79,91	83,58	76,80	0,59	0,80
	SVM POLY	80,00	83,96	76,64	0,60	0,80
Xception Com seleção 135 features	XGBoost	76,28	78,11	74,72	0,53	0,76
	RF	77,66	80,37	76,96	0,55	0,85
	SVM LINEAR	78,44	82,26	75,20	0,56	0,78
	SVM RBF	76,96	85,47	69,76	0,54	0,77
Com seleção 135 features	SVM POLY	79,22	84,52	74,72	0,58	0,79
	XGBoost	75,50	77,74	73,60	0,55	0,76

TABELA V
RESULTADOS DE TESTES NA REDE VGG16

Rede	Classificador	A(%)	S(%)	E(%)	K	AUC
VGG16 Sem seleção 4096 features	RF	78,44	85,09	72,80	0,57	0,85
	SVM LINEAR	78,70	79,05	77,92	0,57	0,78
	SVM RBF	79,65	86,98	73,44	0,59	0,80
	SVM POLY	79,91	88,49	72,64	0,60	0,80
VGG16 Com seleção 120 features	XGBoost	74,81	78,30	71,84	0,50	0,75
	RF	77,57	83,96	72,16	0,55	0,84
	SVM LINEAR	78,78	87,73	71,20	0,57	0,79
	SVM RBF	76,27	89,24	65,28	0,53	0,77
Com seleção 120 features	SVM POLY	78,87	90,00	69,44	0,58	0,79
	XGBoost	74,72	80,38	69,92	0,50	0,75

TABELA VI
RESULTADOS DE TESTES NA REDE VGG19

Rede	Classificador	A(%)	S(%)	E(%)	K	AUC
VGG19 Sem seleção 4096 features	RF	78,18	83,39	73,76	0,56	0,85
	SVM LINEAR	75,75	75,09	76,32	0,51	0,75
	SVM RBF	79,48	88,22	73,76	0,59	0,80
	SVM POLY	79,13	86,79	72,64	0,58	0,79
VGG19 Com seleção 113 features	XGBoost	76,10	79,06	73,60	0,52	0,76
	RF	79,30	85,47	74,08	0,58	0,84
	SVM LINEAR	77,57	83,96	72,16	0,55	0,78
	SVM RBF	76,36	83,01	70,72	0,53	0,76
Com seleção 113 features	SVM POLY	78,00	86,98	70,40	0,56	0,78
	XGBoost	74,89	79,25	71,20	0,50	0,75

(K) e área sob a curva ROC (AUC). Os melhores resultados obtidos em cada método estão destacados em negrito.

Nesse trabalho foram empregados dois cenários de testes.

TABELA VII
RESULTADOS DE TESTES COM TODAS AS FEATURES CONCATENADAS

Rede	Classificador	A(%)	S(%)	E(%)	K	AUC
Concatenadas Sem seleção 12288 features	RF	78,78	81,88	76,16	0,57	0,85
	SVM LINEAR	77,48	76,41	78,40	0,54	0,77
	SVM RBF	80,80	80,62	79,42	0,60	0,80
	SVM POLY	81,29	85,47	77,76	0,62	0,81
Concatenadas Com seleção 244 features	XGBoost	80,87	82,83	79,20	0,62	0,81
	RF	80,51	84,15	77,44	0,61	0,79
	SVM LINEAR	80,51	84,71	76,96	0,61	0,80
	SVM RBF	80,34	86,60	75,04	0,60	0,80
Com seleção 244 features	SVM POLY	81,38	86,60	76,96	0,62	0,81
	XGBoost	81,13	82,83	79,68	0,62	0,81

No primeiro cenário, temos os resultados com todas as características dos modelos. Já no segundo cenário, temos os resultados com a seleção de atributos feitas com Greedy StepWise. No primeiro cenário, foi obtido o melhor resultado com a ResNet50 com classificador SVM Polynomial, com acurácia de 81.73%, sensibilidade de 85.66%, especificidade de 78.40%, Kappa de 0.63 e curva ROC de 0,82. Já no segundo cenário, com seleção de atributos, houve uma queda nos resultados, sendo o melhor com a VGG19 com classificador Randon Forest, com com acurácia de 79.30%, sensibilidade de 85.47%, especificidade de 74.08%, Kappa de 0.58 e curva ROC de 0,84.

Tendo em vista os testes realizados, nota-se que o modelo ResNet50 obteve melhores resultados na maioria dos casos, assim como o classificador SVM com kernel Polynomial, que obteve melhores resultados que os demais classificadores. Os demais resultados obtidos foram satisfatórios de acordo com as métricas especificadas, indicando uma boa capacidade discriminativa e classificatória do algoritmo para as classes definidas.

A comparação com outros trabalhos dentro da literatura é uma tarefa um pouco complicada, já que há algumas diferenças nas maneiras de abordagem do problema e também nas ferramentas e técnicas utilizadas para solucioná-los. A Tabela VIII apresenta um resumo quantitativo em relação aos trabalhos apresentados na Seção II.

TABELA VIII
TABELA DE COMPARAÇÃO DOS RESULTADOS

Trabalho	A(%)	S(%)	E(%)	K	AUC
[4]	93,80	96,50	88,00	-	97,3
[5]	80,00	-	-	-	-
[6]	90,18	-	-	-	-
[7]	-	-	-	-	0,82
Esse trabalho	81,73	85,66	78,40	0,63	0,82

Por fim, o método mostra-se promissor, apesar de não alcançar resultados que excedam o estado da arte, foram aplicadas diferentes estratégias e arquiteturas que possibilitam estudos futuros com base nas experiências apresentadas em neste trabalho.

V. CONCLUSÃO

Neste trabalho foi avaliado e comparado o desempenho de quatro arquiteturas de redes utilizando Transfer Learning para a extração de características e três classificadores para

a classificação automática de lesões de mama em imagens médicas.

A arquitetura ResNet50 atingiu o melhor resultado dentre os algoritmos usados neste trabalho, alcançando valores acima de 78% nas métricas de avaliação, com a utilização do classificador SVM com kernel Polynomial. Os bons resultados na avaliação fornecem oportunidades para propor novas metas de pesquisa na área de lesões de mama.

As arquiteturas de redes investigadas transferiram com sucesso o conhecimento do ImageNet codificados como características convolucionais para o problema na presença de dados limitados de treinamento.

O trabalho apresentado mostra uma boa classificação no uso da Transfer Learning evidenciando uma eficiente ferramenta para a análise automática de imagens, mesmo com um número limitado de imagens e recursos computacionais escassos.

Desta forma, este trabalho apresenta contribuições em duas áreas de estudo. Na área da saúde, oferecendo uma metodologia como segunda opinião a profissionais da área médica para auxiliar no diagnóstico de lesões de mama, e na área da computação, fornecendo informações importantes para a área de processamento de imagens, como forma de auxiliar pesquisas futuras e também de terceiros.

Como trabalhos futuros, pretende-se:

- 1) Utilizar outras arquiteturas de CNN como Transfer Learning para comparação de resultados;
- 2) Utilizar outros classificadores; e,
- 3) Usar métodos automáticos para estimar parâmetros para classificação, afim de melhora os resultados obtidos.

REFERÊNCIAS

- [1] C. B. R. Ferreira and D. L. Borges, "Analysis of mammogram classification using a wavelet transform decomposition," *Pattern Recognition Letters*, vol. 24, no. 7, pp. 973–982, 2003.
- [2] M. S. Tramonte, P. C. Silva, S. R. Chubaci, C. C. Cordoba, G. Zucca-Matthes, and R. A. Vieira, "Delay in diagnosis of breast cancer in a public oncologic hospital," *Medicina (Ribeirão Preto. Online)*, vol. 49, no. 5, pp. 451–462, 2016.
- [3] J. Mezirow, *Transformative dimensions of adult learning*. ERIC, 1991.
- [4] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 737–744.
- [5] Y. Guo, H. Dong, F. Song, C. Zhu, and J. Liu, "Breast cancer histology image classification based on deep neural networks," in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 827–836.
- [6] E. D. Carvalho, A. O. de Carvalho Filho, A. D. de Sousa, A. C. Silva, and M. Gattass, "Method of differentiation of benign and malignant masses in digital mammograms using texture analysis based on phylogenetic diversity," *Computers & Electrical Engineering*, vol. 67, pp. 210 – 222, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0045790617334663>
- [7] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 248 – 257, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0169260715300110>
- [8] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The digital database for screening mammography," in *Proceedings of the 5th international workshop on digital mammography*. Medical Physics Publishing, 2000, pp. 212–218.
- [9] C. C. dos Santos, L. de Assis da Silva, F. M. de Queiroz, R. Picoreti, J. L. A. Samatelo, and R. F. Vassallo, "Uso de transfer learning para o reconhecimento de gestos dinâmicos," in *Congresso Brasileiro de Automática CBA2018*, 2018.

- [10] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [11] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI Global, 2010, pp. 242–264.
- [12] X. Du, M. El-Khamy, J. Lee, and L. Davis, "Fused dnn: A deep neural network fusion approach to fast and robust pedestrian detection," in *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2017, pp. 953–961.
- [13] Z. Chen, Z. Xie, W. Zhang, and X. Xu, "Resnet and model fusion for automatic spoofing detection," in *INTERSPEECH*, 2017, pp. 102–106.
- [14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [16] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.
- [17] E. Azevedo and A. Conci, *Computação gráfica: teoria e prática*. Elsevier, 2008.
- [18] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.
- [19] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001.
- [20] T. Chen, T. He, M. Benesty, V. Khotilovich, and Y. Tang, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, pp. 1–4, 2015.
- [21] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [22] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *The Annals of Statistics*, vol. 28, pp. 337–407, 04 2000.
- [23] R. Song, S. Chen, B. Deng, and L. Li, "extreme gradient boosting for identifying individual users across different digital devices," in *Web-Age Information Management*, B. Cui, N. Zhang, J. Xu, X. Lian, and D. Liu, Eds. Cham: Springer International Publishing, 2016, pp. 43–54.
- [24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [25] R. Caruana and D. Freitag, "Greedy attribute selection," in *Machine Learning Proceedings 1994*. Elsevier, 1994, pp. 28–36.



Rafael Pedro da Silva Neto atualmente aluno de graduação do curso de Bacharelado em Sistema de Informação na Universidade Federal do Piauí (UFPI), Brasil. Seus principais interesses de pesquisa incluem abordagens de *Deep Learning* e desenvolvimento de aplicações móveis.



Antonio Oseas de Carvalho Filho recebeu um Ph.D. em Engenharia Elétrica pela Universidade Federal do Maranhão - Brasil em 2016. Atualmente, é professor da Universidade Federal do Piauí (UFPI).