COMPARAÇÃO DE ARQUITETURAS DE REDES NEURAIS ARTIFICIAIS CONVOLUCIONAIS PARA CLASSIFICAÇÃO DE TOMOSSÍNTESES.

Gabriel Carvalho Santana[1] (PROITI)

Guilherme Apolinário Silva Novaes[2] (Orientador)

Universidade Católica de Santos

Curso: Ciência da Computação

[1]gabriel.carvalho@unisantos.br; [2]gabriellccarvalho13@gmail.com; [2]g.novaes@unisantos.br

RESUMO:

Objetivando a criação de ferramentas tecnológicas que agilizem o processo de análise de tomossínteses (mamografias 3D), o *dataset Digital Breast Tomosynthesis (Breast-Cancer-Screening-DBT)* foi utilizado e submetido a um conjunto de arquiteturas de redes neurais artificiais (ResNET-18, ResNET-34, SE ResNET-18, SE ResNET-34, VGG-16 , GoogLeNET e Inception-V3.) para classificar as imagens, visando tanto diferenciar uma classe específica das demais, quanto classificar as imagens entre todas as possibilidades (*Normal, Actionable, Benign e Cancer*). As imagens, que originalmente são tridimensionais, foram tratadas como imagens de duas dimensões com um único filtro de pixels para os treinos e para as predições. No momento da predição, a resposta é dada com base na soma das probabilidades obtidas por cada uma das fatias que compõem a imagem. A acurácia nos dados de treino mostrou que as redes utilizadas são promissoras. Para as classificações binárias, a arquitetura GoogLeNET se apresentou como a melhor para classificar imagens da classe *Normal* das demais (84,44% de acurácia total), enquanto a SE ResNET-18 se mostrou a melhor escolha para diferenciar a classe *Actionable* do restante (94,44% de acurácia), para a classe *Benign* o melhor algoritmo foi a Inception-V3 (93,33% de acurácia) e pôr fim a arquitetura que mais se destacou em diferenciar a classe *Cancer* das demais foi a novamente a SE ResNET-18 (91,11% de acurácia). Para a classificação múltipla, todas tiveram resultados próximos, porém a GoogLeNET foi a melhor, tendo 86,66% de acurácia total nos testes, enquanto a ResNET-18 e ResNET-34 obtiveram os piores resultados (77,77% de acurácia em ambas).

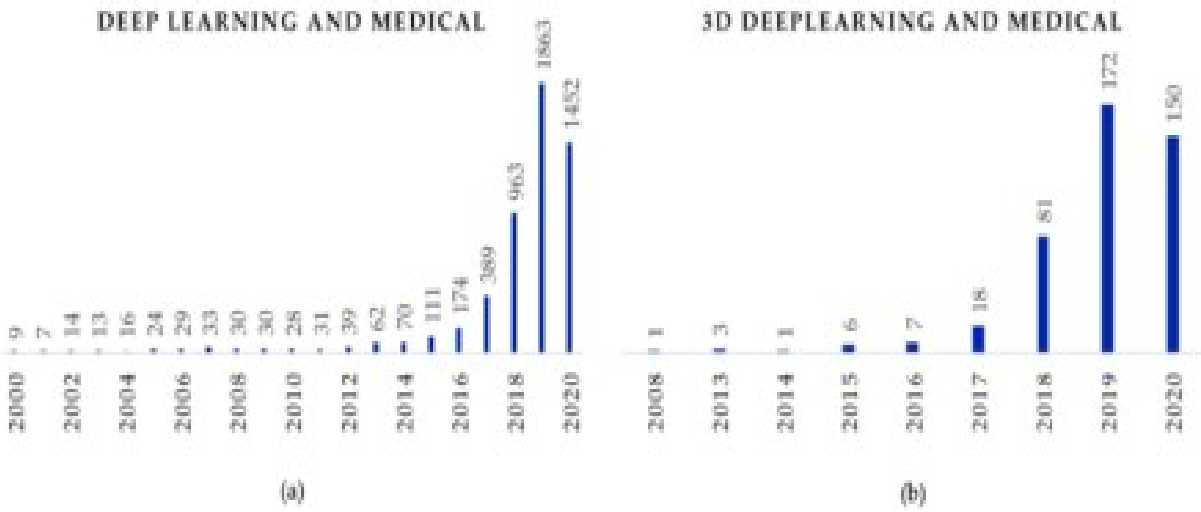
Palavras-chave: Processamento de imagens médicas, Processamento de Tomossínteses, Classificação de Imagens.

1 INTRODUÇÃO

Desde o advento da criação da primeira arquitetura com redes convolucionais conhecida, chamada AlexNET (KRIZHEVSKY *et al.*, 2012), diversas novas arquiteturas foram surgindo, além de diversos campeonatos como LVIS (*Large Vocabulary Instance Segmentation*) e ILSVRC (*ImageNet Large Scale Visual Recognition Challenge, criado por* Olga Russakovsky, *et al.*), que fomentaram o desenvolvimento de novas técnicas para o campo de visão computacional. Com isso, não demorou para se começar a usar tais tecnologias na área

médica, além de adaptar algoritmos para tridimensionalidade e haver campeonatos específicos nessa área como o Medical Segmentation Decathlon (ANTONELLI *et al.*, 2018), em que o objetivo era segmentar imagens tridimensionais de ressonâncias magnéticas em diversos órgãos humanos. Na figura 1 é possível ver o aumento crescente de trabalhos que utilizaram técnicas de *Deep Learning* em artigos na plataforma PubMed.

Figura 1 – Uso de *Deep Learning* em pesquisas médicas encontradas na PubMed



Fonte: Singh, Satya, *et al.* (2020)

Considerando o contexto vivido atualmente, com grandes evoluções em hardware, software e dados abertos que possibilitam pesquisas em diversas áreas, o trabalho se propõe a explorar alguns dos principais algoritmos conhecidos de visão computacional na base de dados *Digital Breast Tomosynthesis (Breast-Cancer-Screening-DBT)*, com o intuito de realizar comparações entre eles e facilitar a escolha para trabalhos futuros de outros pesquisadores ou a continuação deste.

2. PROCEDIMENTOS DE PESQUISA

2.1 Objetivos da pesquisa

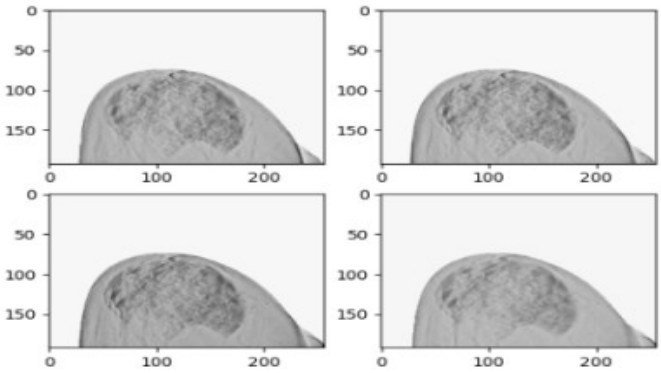
A pesquisa tem como objetivo estudar um conjunto de dados que possui imagens de tomossínteses e testar arquiteturas de redes neurais artificiais com camadas convolucionais que possam processar e classificar tal tipo de dado para detectar a existência ou não de anormalidades.

2.2 Dados utilizados

Os dados pertencem ao conjunto de dados *Breast Cancer Screening - Digital Breast Tomosynthesis* (BUDA *et al.*, 2022) o qual possui imagens de tomossínteses de mamas femininas no formato de arquivo DICOM (comumente utilizado em radiologia). As imagens (221 no total) estão divididas em quatro classes distintas: *Normal* (35% do total), *Actionable* (25% do total), *Benign* (23% do total) e *Cancer* (17% do total). Somado a isso, são dados que formam tensores de quarta ordem e não possuem padrão de tamanho na primeira e terceira dimensão, além de possuírem diferentes orientações visuais, os únicos padrões são em relação ao valor dos pixels que variam entre 0 e 1023 com apenas um canal de cor e a segunda

dimensão que é um valor fixo. A figura 2 mostra alguns exemplos de uma das mamas dos dados com variação na terceira dimensão.

Figura 2 - Exemplos de diferentes camadas de uma imagem extraídas do *dataset*



Fonte: Buda *et al.* (2022)

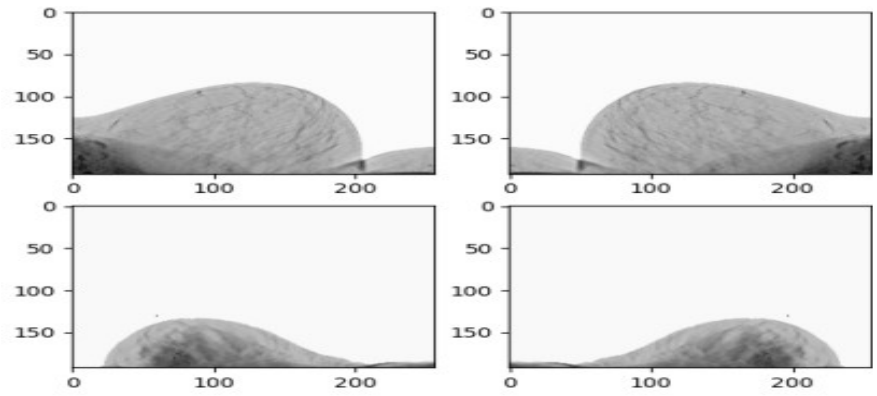
2.3 Preparação dos dados

Para viabilizar a utilização dos dados em algoritmos de Inteligência Artificial foram necessários processamento e normalização prévias, de modo a fixá-los em um padrão de tamanho e tipo predefinidos. Para isso, os dados foram convertidos do formato DICOM para npy, e em seguida processados para apresentar dimensões fixas de 192 x 256 x profundidade x 1, em que a profundidade é a terceira dimensão da imagem e representa quantas imagens bidimensionais formam a imagem tridimensional. Originalmente as dimensões das imagens variam entre 1830 e 1960 na primeira dimensão, fixas em 2457 na segunda dimensão, 27 a 106 camadas na terceira dimensão e 1 na última dimensão (representando um canal de 10 bits com 1024 possibilidades de tons de cinza). A terceira camada não foi alterada, pois as sub imagens de cada imagem original foi tratado individualmente no algoritmo e utilizadas em conjunto para as predições finais. Além disso, os dados foram normalizados em uma escala de 0 a 1 ao invés de 0 a 1023.

2.3.1 Processo de *Data Augmentation*

Na fase de treinamento, para aumentar as possibilidades de encontro de padrões nos dados, foi utilizado o processo de espelhamento das imagens. Na figura 3 há dois exemplos.

Figura 3 – exemplos do processo de *Data Augmentation* empregado



Fonte: Autor

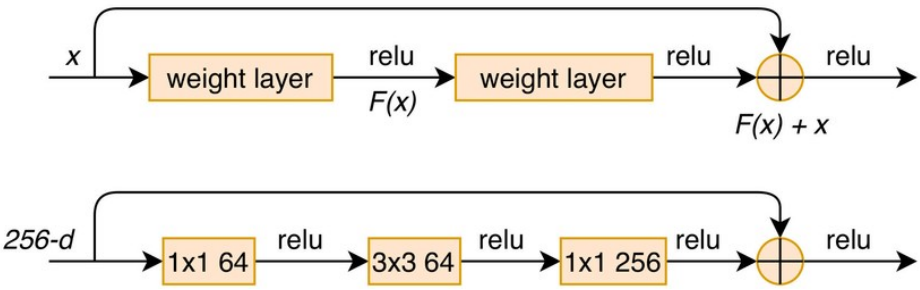
2.4 Arquiteturas utilizadas

Para o desenvolvimento da pesquisa, foram escolhidas 7 arquiteturas de redes neurais artificiais convolucionais e todas foram submetidas a testes para diferenciar cada uma das classes das demais, quanto classificar a imagem dentre as quatro *labels* possíveis. As Redes Neurais Artificiais escolhidas foram: ResNET-18, ResNET-34, SE ResNET-18, SE Resnet-34, VGG-16, GoogLeNET (Inception-V1) e Inception-V3.

2.4.1 Arquiteturas residuais

As ResNETs (Redes Neurais Residuais) é uma família de arquiteturas de redes neurais convolucionais proposta por Kaiming He, Xiangyu Zhang, Shaoqing Ren e Jian Sun em 2015. Ela foi projetada para resolver o problema do desaparecimento do gradiente *(Vanishing Gradient)*, que ocorre quando as redes neurais ficam muito profundas e não conseguem aprender, pois as derivadas parciais dos pesos tendem a zero. Isso torna difícil para a rede aprender e pode levar a um desempenho ruim. As ResNETs resolvem o problema do *Vanishing Gradient* usando conexões residuais, que são caminhos diretos que permitem que a saída de uma camada seja adicionada à saída da camada seguinte. Isso permite à rede "pular" algumas camadas e ainda assim aprender. A proposta revolucionou a área de visão computacional, permitindo a criação de arquiteturas maiores do que as já existentes e que ainda possuem capacidade de encontrar padrões nos dados. Na figura 4 é demonstrado como são os blocos residuais.

Figura 4: Blocos Residuais



Fonte: Eftim Zdravevski (2020)

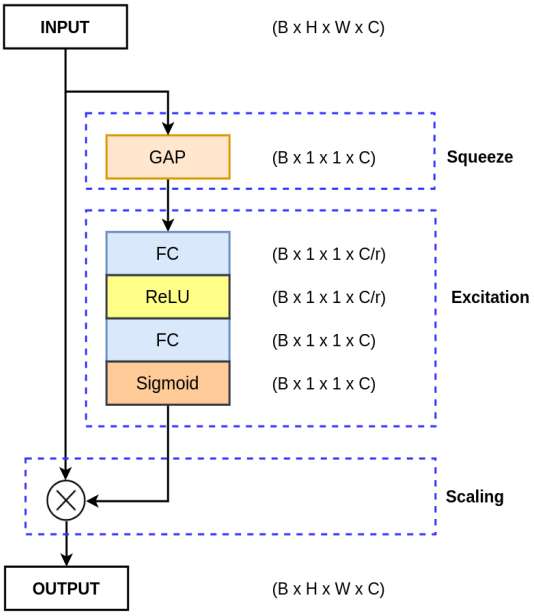
2.4.2 Variação Squeeze and Excitation para ResNETs

As SE ResNETs (Squeeze-and-Excitation Residual Networks) é uma variante das ResNETs originais, proposta por Kaiming He, Xiangyu Zhang, Shaoqing Ren e Jian Sun, que incorpora blocos Squeeze-and-Excitation (SE) para permitir que a rede realize uma recalibração dinâmica dos filtros de convolução durante a fase de treino do algoritmo.

O bloco SE é composto por três partes, *Squeeze*: um processo de G*lobal Average Pooling (GAP)*, usado para reduzir a dimensão do tensor de entrada para uma única dimensão, *Excitation*: uma rede neural com duas camadas densas para calcular um coeficiente de ponderação para cada um dos filtros que compõem o tensor de entrada e *Scaling*: o processo de multiplicação do tensor de entrada pelo coeficiente de ponderação calculado pelo bloco SE.

Os blocos SE permitem que o algoritmo atribui pesos diferentes para cada filtro das camadas de convolução. Isso permite que a rede aprenda representações mais discriminativas dos dados, descartando filtros pouco relevantes. Na figura 5 é mostrado o diagrama de um bloco SE.

Figura 5: Diagrama Squeeze-and-Excitation

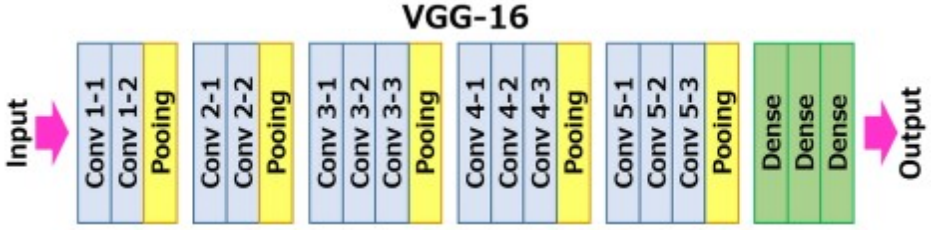


Fonte: Nikhil Tomar (2021)

2.4.3 Arquitetura VGG-16

A VGG-16 é uma arquitetura de rede neural artificial que foi proposta por Karen Simonyan e Andrew Zisserman em 2014. Ela foi projetada para ser uma arquitetura simples e fácil de implementar. A rede é composta por 16 camadas treináveis, além de 4 camadas para redução dos dados conhecidas como *Pooling Layers.* As primeiras 13 camadas são convolucionais e as três últimas são camadas densas totalmente conectadas. A figura 6 ilustra a arquitetura completa.

Figura 6: Arquitetura VGG-16

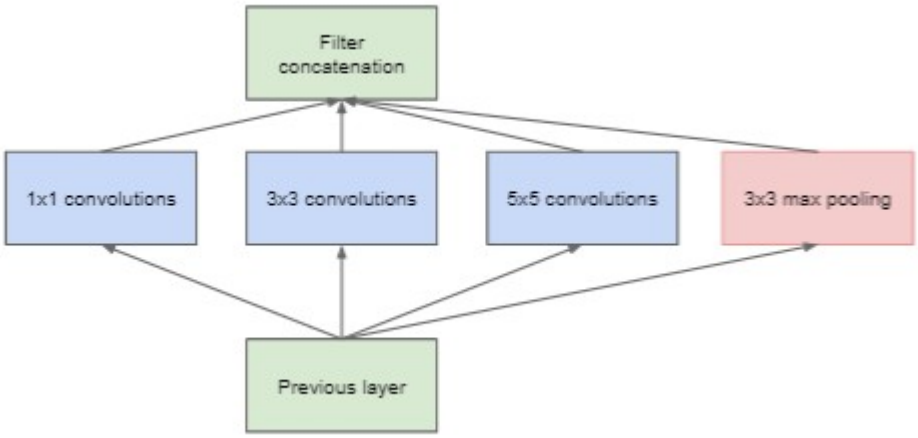


Fonte: Abhay Parashar (2020)

2.4.4 Arquitetura GoogLeNET

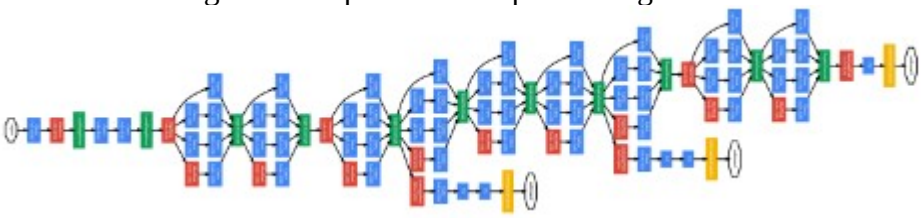
A GoogLeNET (também conhecida como Inception-V1) é uma arquitetura proposta por Christian Szegedy *et al*. em 2014. Ela foi projetada para resolver o problema do aumento da complexidade computacional que grandes arquiteturas traziam, juntamente do *Vanishing Gradient* que ocorre quando as redes neurais ficam muito profundas. A GoogLeNet usa um conceito chamado *Inception module*, que permite que a rede combine diferentes tamanhos de filtros de convolução em um único bloco, isso permite que a rede aprenda representações mais complexas dos dados apresentados, mesmo com um número relativamente pequeno de parâmetros. A figura 7 mostra um bloco *Inception* e a figura 8 a arquitetura completa da GoogLeNET com suas duas saídas auxiliares.

Figura 7: Ilustração de um *Inception Module*



Fonte: Bo Zhao (2017)

Figura 8: Arquitetura completa GoogLeNET.

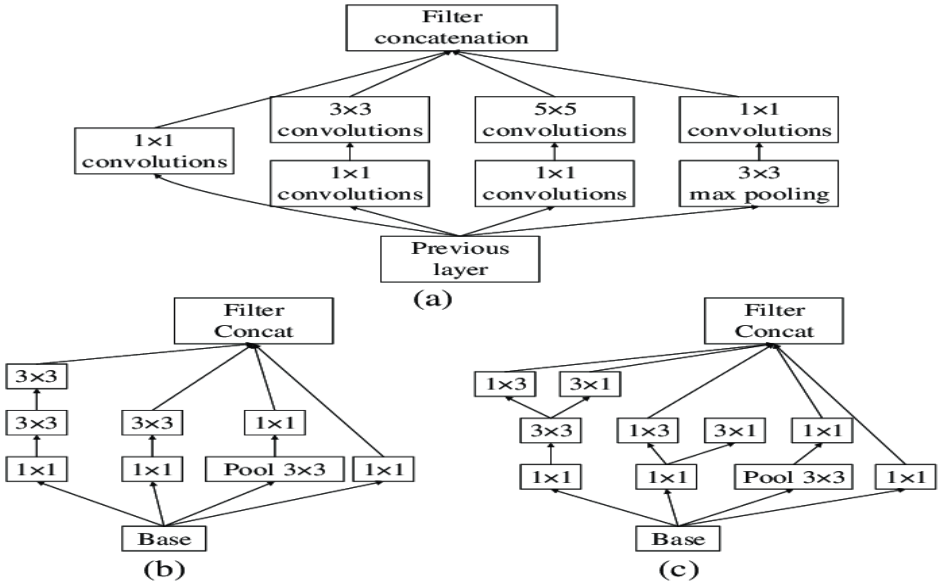


Fonte: Sai Kumar Basaveswara (2017)

2.4.5 Arquitetura Inception-V3

Inception-V3, proposta em 2016 pelo mesmo autor da GoogLeNET. Ela foi projetada para melhorar o desempenho da GoogLeNET na classificação de imagens e detecção de objetos. Entre suas principais diferenças, são utilizados três tipos de *Inception Module* diferentes do original, possuindo maiores variações em quantidade e número de parâmetros do que o da GoogLeNET. Na figura 9 é mostrado os novos blocos que compõem a arquitetura.

Figura 9: Variações do *Inception module*



Fonte: Zhiyu Qu, et al (2019)

2.5 Comparação entre o tamanho dos modelos

Além da forma como as arquiteturas se apresentam, também variam em número de camadas - tanto treináveis (Densas e Convolucionais), quanto não treináveis (*Poolings, Dropout e Batch Normalization)* - e quantidade de parâmetros. A tabela 1 mostra a comparação entre elas.

Tabela 1 - Comparação entre números de camadas e parâmetros entre os modelos utilizados

Nome do modelo	Quantidade de camadas treináveis	Quantidade de parâmetros totais
ResNET-18	18	14.843.073
ResNET-34	34	28.325.569
SE ResNET-18	45	1.892.644
SE ResNET-34	65	28.959.025
VGG-16	16	10.172.228
GoogLeNET	47	5.305.636
Inception-V3	106	15.664.340

Fonte: Autor

2.6 Divisão dos dados, processos de treinamento e predição

Inicialmente, cada uma das arquiteturas mencionadas passou por testes em dois tipos de problemas distintos: diferenciar uma classe específica das demais e classificar cada imagem em uma das quatro categorias possíveis. Posteriormente, os dados foram divididos de maneira estratificada, com 80% destinados ao treinamento e 20% para fins de teste. Durante o processo de treinamento, realizamos cinco divisões, dividindo os dados de treinamento em 68% para o treinamento efetivo e 12% para fins de validação. Essas divisões foram feitas de forma aleatória, mas mantendo a estratificação em cada uma delas, visando avaliar a consistência dos modelos.

Para cada arquitetura e objetivo, selecionamos o conjunto de pesos que apresentou a melhor métrica ROC-AUC (Receiver Operating Characteristic Curve - Área Sob a Curva) ao longo das cinco divisões e, em seguida, realizamos um treinamento adicional de 5 épocas, utilizando todos os dados de treinamento disponíveis. Esses modelos treinados foram posteriormente utilizados para fazer previsões nos dados de teste.

Durante o processo de treinamento, os resultados obtidos estavam relacionados às probabilidades de acerto para cada sub-imagem individual. No teste final, agregamos as probabilidades de todas as sub-imagens para prever o resultado completo de cada imagem.

2.7 Código-fonte

Os códigos do projeto com a implementação de todas as arquiteturas utilizadas, processamento de dados e treinamentos encontram-se em um repositório aberto na plataforma GitHub[1].

3. RESULTADOS E DISCUSSÃO

Após realizar todos os treinos e validações, todas as arquiteturas foram utilizadas para predizer o conjunto de imagens que foi guardado para teste. A tabela 1 mostra os resultados de cada um dos algoritmos com a acurácia de cada classe e a acurácia total. A figura 2 faz o mesmo, porém mostra a acurácia da classe objetivo, acurácia do restante das classes previstas

[1] Link para o repositório https://github.com/obb199/Breast-Cancer-Screening-DBT_Classification

em conjunto e pôr fim a acurácia total, além disso, a figura 10 mostra as curvas ROC e o valor de AUC dos melhores classificadores binários e a figura 11 mostra a matriz de confusão do melhor classificador múltiplo (GoogLeNET).

Tabela 2: Acurácia para os casos binários

MODELO	CLASSE DE DIFERENCIAÇÃO	ACURÁCIA NAS CLASSES GERAIS	ACURÁCIA NA CLASSE OBJETIVO	ACURÁCIA TOTAL
ResNET-18	Normal	87,50%	68,96%	75,5%
ResNET-18	Actionable	100,00%	21,21%	42,22%
ResNET-18	Benign	81,81%	85,29%	84,44%
ResNET-18	Cancer	16,66%	94,87%	84,44%
SE ResNET-18	Normal	75,00%	75,86%	75,55%
SE ResNET-18	Actionable	83,33%	100,00%	95,55%
SE ResNET-18	Benign	63,63%	97,05%	88,88%
SE ResNET-18	Cancer	50,00%	97,43%	91,11%
ResNET-34	Normal	75,00%	82,75%	80,00%
ResNET-34	Actionable	91,66%	96,96%	95,55%
ResNET-34	Benign	63,63%	97,05%	88,88%
ResNET-34	Cancer	97,43%	33,33%	88,88%
SE ResNET-34	Normal	60,60%	25,00%	51,11%
SE ResNET-34	Actionable	96,96%	75,00%	91,11%
SE ResNET-34	Benign	85,29%	81,81%	84,44%
SE ResNET-34	Cancer	92,30%	50,00%	86,66%
GoogLeNET	Normal	75,86%	100,00%	84,44%
GoogLeNET	Actionable	100,00%	66,66%	91,11%
GoogLeNET	Benign	55,88%	81,11%	62,22%
GoogLeNET	Cancer	94,87%	50,00%	88,88%
Inception-V3	Normal	100,00%	56,25%	84,44%
Inception-V3	Actionable	96,96%	75,00%	91,11%
Inception-V3	Benign	100,00%	72,72%	93,33%
Inception-V3	Cancer	100,00%	33,33%	91,11%
VGG-16	Normal	72,41%	37,50%	60,00%

VGG-16	Actionable	96,96%	75,00%	91,11%
VGG-16	Benign	97,05%	63,63%	88,88%
VGG-16	Cancer	0,0%	100,00%	13,33%

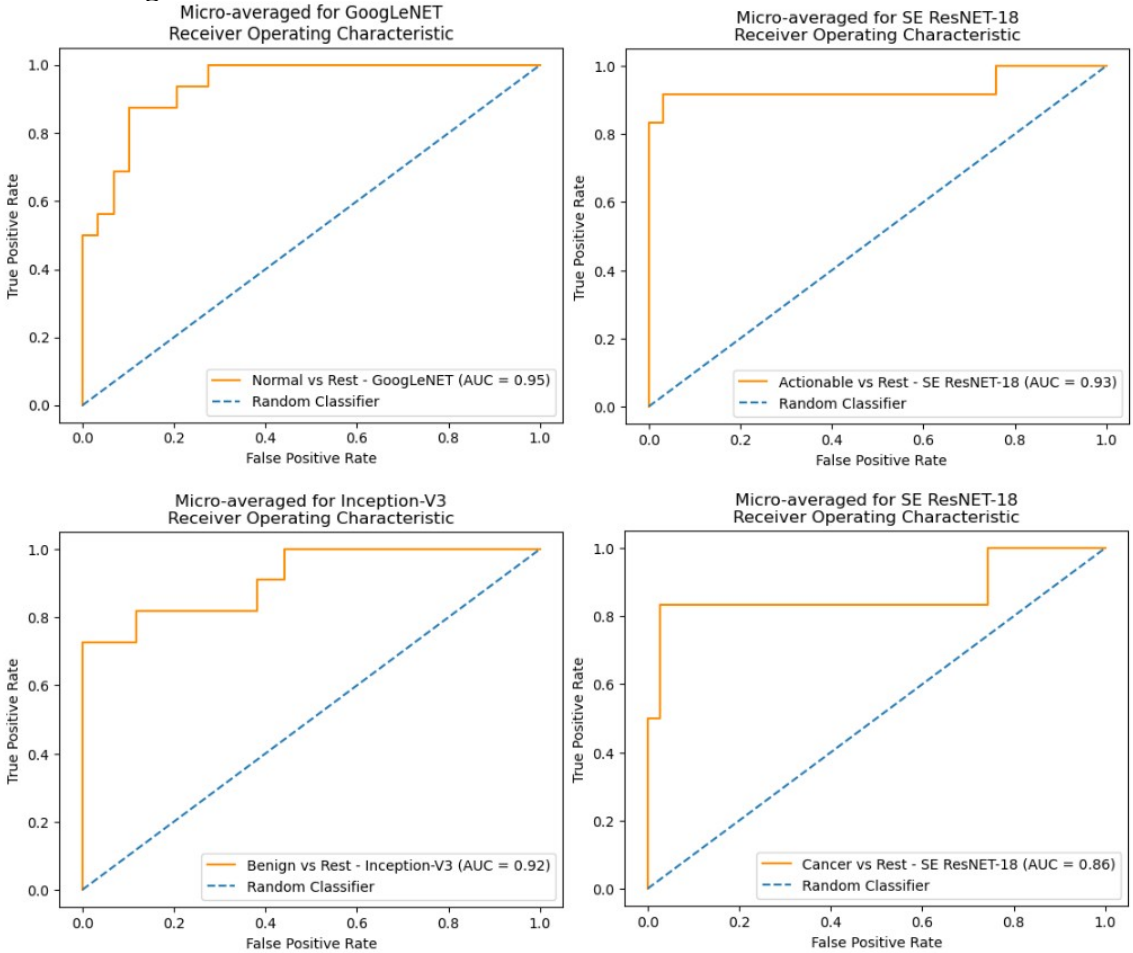
Fonte: Autor

Tabela 3: Acurácia para os casos de múltiplos rótulos.

MODELO	ACURÁCIA NORMAL	ACURÁCIA ACTIONABLE	ACURÁCIA BENIGN	ACURÁCIA CÂNCER	ACURÁCIA TOTAL
ResNET-18	75,00%	83,33%	81,81%	66,66%	77,77%
SE ResNET-18	81,25%	83,33%	72,72%	83,33%	80,00%
ResNET-34	81,25%	83,33%	72,72%	66,66%	77,77%
SE ResNET-34	81,25%	91,66%	72,72%	66,66%	80,00%
GoogLeNET	100,00%	91,66%	72,72%	66,66%	86,66%
Inception-V3	81,25%	91,66%	72,72%	66,66%	80,00%
VGG-16	75,00%	91,66%	72,72%	66,66%	81,81%

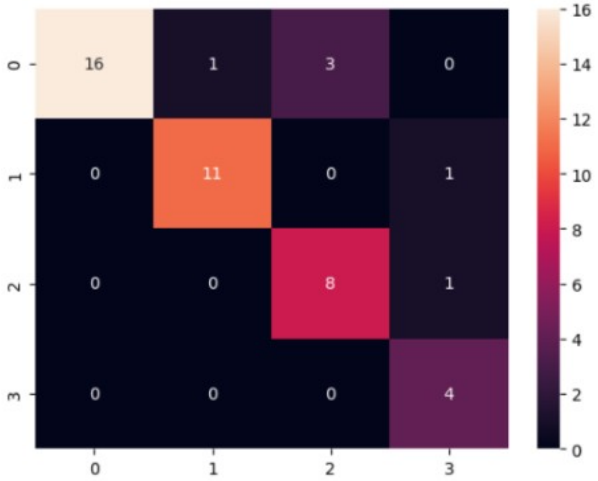
Fonte: Autor

Figura 10 - Curvas ROC e valor de AUC dos melhores classificadores



Fonte: Autor\

Figura 11 - Matriz de confusão do melhor classificador múltiplo



Fonte: Autor

4 CONSIDERAÇÕES FINAIS

A técnica de treinar o algoritmo para prever cada fatia das imagens e no momento das predições utilizar a soma das probabilidades dadas para cada uma das sub imagens se mostrou muito eficiente, o que pode ser um caminho para o desenvolvimento de novos algoritmos de *Machine Learning* na área de classificação de imagens médicas volumétricas. Somado a isso, *o* método mostra-se computacionalmente econômico por não precisar processar toda a imagem de uma só vez.

5 REFERÊNCIAS

BUDA, Saha; *et al*. (2020). **Breast Cancer Screening – Digital Breast Tomosynthesis (Breast-Cancer-Screening-DBT)** [Data set]. The Cancer Imaging Archive.

GUPTA, Agrim; *et al*. **LVIS: A Dataset for Large Vocabulary Instance Segmentation.** 2019.

HE, Kaiming; ZHANG, Xiangyu; REN Shaoqing; SUN, Jia. **Deep Residual Learning for Image Recognition**. Microsoft Research. 2015.

HU, Jie; SHEN, Li; ALBANIE, Samuel; SUN, Gang; WU, Enhua. **Squeeze-and-Excitation Networks**. Momenta e Oxford University. 2016.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey. **ImageNet Classification with Deep Convolutional Neural Networks**. NIPS, 2012.

PETROVSKA, Biserka; *et al*.. **Aerial Scene Classification through Fine-Tuning with Adaptive Learning Rates and Label Smoothing**. MDPI, 2020.

REINKE, Antonelli; *et al*. **The Medical Segmentation Decathlon**. Nat Commun 13, 4128, 2022.

SIMONYAN, Karen e ZISSERMAN, Andrew. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. Visual Geometry Group, Department of Engineering Science, University of Oxford. 2015.

RUSSAKOVSKY, Olga; et al. **ImageNet Large Scale Visual Recognition Challenge.** 2014.

SZEGEDY, Christian; *et al*. **Going deeper with convolutions**. Google Inc, University of North Carolina, Chapel Hill e University of Michigan. 2014.

SZEGEDY, Christian; *et al*. **Rethinking the Inception Architecture for Computer Vision**. Zbigniew Wojna University College London. 2015.