

# **AValiação DE DESEMPENHO DE REDES NEURais CONVOLUCIONais PARA DETECÇÃO DE ANOMALIAS EM IMAGENS HISTOPATOLÓGICAS DE CâNCER DE MAMA**

## **PERFORMANCE EVALUATION OF CONVOLUTIONAL NEURAL NETWORKS FOR DETECTING ANOMALIES IN HISTOPATHOLOGICAL IMAGES OF BREAST CANCER**

Robson de Sousa dos Santos\*

Silas Santiago Lopes Pereira\*\*

Raimundo Valter Costa Filho\*\*\*

### **RESUMO**

Diversas abordagens têm sido exploradas na literatura de engenharia biomédica para a análise e interpretação automatizada de imagens clínicas. Desafios existentes incluem: o tratamento do grande volume de dados de imagens médicas, a variação na coloração nas representações, a qualidade e a disponibilidade de informações de diferentes categorias de doenças. Técnicas de visão computacional podem ser aplicadas no apoio às diferentes tarefas no campo da análise de imagens médicas. Abordagens de aprendizagem profunda têm alcançado resultados promissores. Este trabalho propõe explorar a aplicação de técnicas de aprendizado de máquina, mais especificamente aprendizado profundo, em problemas de detecção de anomalias em imagens histopatológicas de câncer de mama. O desempenho de várias arquiteturas baseadas em redes neurais convolucionais foram avaliadas. O modelo atingiu uma acurácia balanceada de 82,24% e *F-score* de 81,58%. Os resultados obtidos foram comparáveis e competitivos quando comparados com o estado da arte.

**Palavras-chave:** Sistemas de apoio de decisão. Diagnóstico por imagem. Aprendizado profundo.

### **ABSTRACT**

---

\* Robson Santos é graduando em ciência da computação no Instituto Federal do Ceará (IFCE) e pesquisador bolsista pelo Programa Institucional de Bolsas de Iniciação Científica (PIBIC). Endereço eletrônico: allrobson99@gmail.com.

\*\* Silas Santiago é professor do IFCE, mestre em ciências da computação na universidade estadual do ceará (UECE) e bacharel em ciência da computação pela UECE. Endereço eletrônico: silas@lar.ifce.edu.br.

\*\*\* Raimundo Valter é doutor em Engenharia de Teleinformática (UFC), mestre em Infraestrutura de Redes (UFERSA/UERN), especialista em Docência no Ensino Superior (IBF) e graduado em ciência da computação (UFERSA). Atualmente é professor do IFCE Aracati, desenvolvendo pesquisas na área de aprendizado de máquina aplicado à saúde e sistemas embarcados.. Endereço eletrônico: valter.costa@ifce.edu.br.

Several approaches have been explored in the biomedical engineering literature for automated analysis and interpretation of clinical images. Existing challenges include: handling the large volume of medical imaging data, the variation in coloration in the representations, the quality and availability of information from different categories of diseases. Computer vision techniques can be applied to support different tasks in the field of medical image analysis. Deep learning approaches have achieved promising results. This work proposes to explore the application of machine learning techniques, more specifically deep learning, in problems of detecting anomalies in histopathological images of breast cancer. The performance of several architectures based on convolutional neural networks were evaluated. The model achieved a balanced accuracy of 82.24% and *F-score* of 81.58%. The results obtained were comparable and competitive when compared to the state of the art.

**Keywords:** Deep Learning. Diagnostic imaging. Decision support systems.

## 1 INTRODUÇÃO

Nos últimos anos os hospitais têm produzido imensa quantidade de diagnósticos baseados em imagens. Essas imagens são de grande importância, pois permitem extrair informações que tornam o tratamento de tumores mais eficaz. Na análise desses dados um olho humano treinado é imprescindível para a interpretação correta e indicação de diagnóstico.

Contudo, pesquisas recentes vêm tentando, através de algoritmos de inteligência artificial, criar mecanismos de detecção automática de anomalias em imagens médicas com objetivo de apoiar o diagnóstico de doenças. Esses algoritmos vêm auxiliando no trabalho desses especialistas, reduzindo o tempo de diagnóstico da doença e a taxa de erros médicos, agilizando o tratamento dos pacientes (ANWAR et al., 2018).

A aprendizagem profunda, promete desempenhar um importante papel nessa categoria de pesquisa, tornando o trabalho de diagnóstico por imagem um processo mais rápido e eficiente para os humanos (CHALAPATHY; CHAWLA, 2019). Entretanto, do ponto de vista das máquinas, o aprendizado profundo pode não ser uma tarefa trivial, dado que há alguns problemas na hora de analisar essas imagens, sendo o caso da variação na coloração, dos diferentes graus da doença e, com a falta de classificadores e modelos consagrados na preparação desses dados, o trabalho do diagnóstico de imagem torna-se desafiador (ANWAR et al., 2018).

O aprendizado profundo, quando aplicado à classificação, é um exemplo de ferramenta do aprendizado de máquina em que a cada passo melhora as representações aprendidas dos dados subjacentes visando atingir ao máximo a separabilidade de classes (ANWAR et al., 2018). As técnicas de aprendizado profundo demandam, além de uma abundância de dados para treinamento, um substancial poder de processamento. Em alguns casos, quando a quantidade de dados de treinamento não é suficiente, métodos como *data augmentation* e *transfer learning* podem ser aplicados (JANOWCZYK; MADABHUSHI, 2016).

Nos últimos anos, os algoritmos de detecção de anomalias baseados no aprendizado

profundo estão se popularizando em um conjunto diversificado de tarefas. Algumas evidências revelam que o aprendizado profundo supera completamente os métodos tradicionais (CHALAPATHY; CHAWLA, 2019). Nesse contexto, o objetivo dessa pesquisa é explorar a aplicação de aprendizado profundo em problemas de detecção de anomalias em imagens histopatológicas de câncer de mama, de modo a utilizar redes neurais convolucionais em classificações binárias para detecção de estruturas visuais que denotem anomalias.

Este trabalho está organizado da seguinte forma: na Seção 2, será apresentada a fundamentação teórica; na Seção 3, os trabalhos relacionados será apresentado e argumentado; na Seção 4, um detalhamento da proposta será mostrada; na Seção 5 será apresentado os resultados esperados da pesquisa científica; por fim, na Seção 6, estará contido as conclusões e os trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção será apresentada a fundamentação teórica necessária para um aprofundamento melhor do entendimento da proposta apresentada. No decorrer deste capítulo será explicado os contextos médicos abrangentes da pesquisa e os principais conceitos relacionados ao processo detecção de anomalias, aprendizado de máquina, aprendizado profundo e *Rede Neural Convolucional* (CNN).

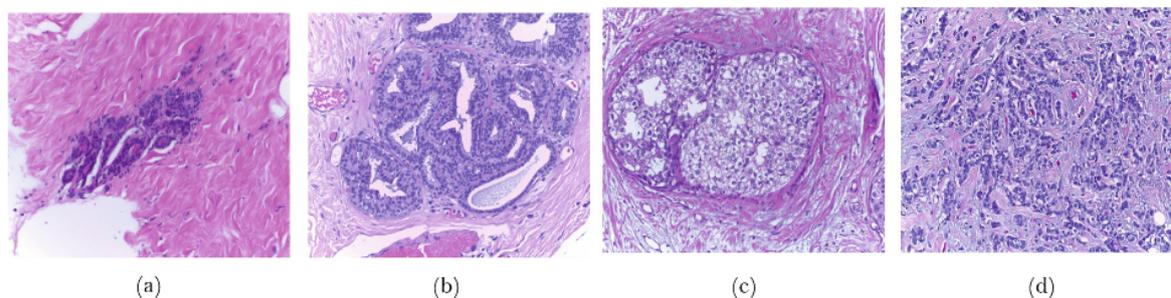
### 2.1 Detecção de Anomalias em Imagens Médicas

Com finalidade de melhorar o processo de análise de imagens médicas muitas tecnologias e técnicas vêm sendo adotadas por pesquisadores e cientistas. Parte importante do processo de melhoria da análise é auxiliar especialistas clínicos e médicos a tornar o processo de diagnóstico e tratamento mais eficiente (CHALAPATHY; CHAWLA, 2019). O diagnóstico precoce e eficaz permite ao médico descrever rapidamente a situação dos pacientes que esperam os resultados desses exames clínicos, desempenhando um papel vital na detecção de informações anatômicas no corpo humano.

*Digital pathology* (DP) ou patologia digital é o processo pelo qual lâminas histológicas são digitalizadas para produzir imagens médicas de alta resolução. Essas imagens digitalizadas oferecem a possibilidade de aplicar técnicas de análise de imagem ao DP para aplicações em detecção e classificação. Uma série de tarefas de análise de imagem em DP envolve alguma categoria de classificação de tecido para extrair características. Como resultado, há uma forte necessidade de desenvolver algoritmos eficientes e robustos para análise de imagens DP (JANOWCZYK; MADABHUSHI, 2016). A Figura 1 mostra as imagens microscópicas de tecidos mamários de não carcinoma (normal e benigno) e carcinoma (In Situ e invasivo), a qual é baseada em (NAWAZ et al., 2018).

A detecção de anomalias em imagens médicas é o processo de identificação de uma determinada categoria de doença, como um tumor, por exemplo. Nessa tarefa, o *software* percorre um conjunto de eventos ou objetos e sinaliza alguns deles como incomuns ou atípicos

Figura 1 – Imagens microscópicas de tecidos mamários, onde (a) e (b) representam tecidos não carcinoma (normal e benigno) e (c) e (d) carcinoma (in situ e invasivo), respectivamente.



Fonte: (NAWAZ et al., 2018)

(GOODFELLOW; BENGIO; COURVILLE, 2016). Tradicionalmente os especialistas clínicos detectam anormalidades, esse processo se dá pela análise de imagens médicas feita manualmente pelo médico ou outro profissional qualificado, tendo este que avaliar sem o auxílio de nenhum programa de inteligência artificial se há a presença ou não de alguma anomalia nas imagens clínicas. Esse processo consome bastante tempo e requer um esforço humano grande (ANWAR et al., 2018).

Ainda que importante, o processo de extrair informações dos conjuntos de dados a partir das imagens clínicas por meio do trabalho manual pode ser uma tarefa demorada e exaustiva (CHALAPATHY; CHAWLA, 2019).

Para que a detecção de anomalias em imagens médicas aconteça, deve o algoritmo produzido conseguir distinguir um dado anômalo de um dado normal. Os rótulos indicam se uma instância de dados escolhida é normal ou um *outlier*. Um *outlier* pode representar uma anomalia, ou seja, são dados que se diferenciam de todos os outros dados rotulados como normais. As anomalias são entidades raras, logo, é um desafio obter seus rótulos. Além disso, o comportamento anômalo pode mudar temporalmente (CHALAPATHY; CHAWLA, 2019).

Os modelos de detecção profunda de anomalias podem ser amplamente classificados em categorias com base na extensão da disponibilidade dos seus rótulos (CHALAPATHY; CHAWLA, 2019). A seguir são apresentadas às três categorias.

- A **detecção profunda e supervisionada de anomalias** aprende o limite de separação de um conjunto de instâncias de dados passados pelo treinamento e, em seguida, classifica uma instância de teste em classes normais ou anômalas a partir do modelo aprendido (CHALAPATHY; CHAWLA, 2019).
- Na **detecção profunda e semi supervisionada de anomalias**, pressupõem que todas as instâncias de treinamento têm apenas um rótulo de classe. A obtenção de rótulos de classes positivas são mais fáceis quando comparadas com a obtenção das categorias que

contêm anomalias, em razão disso, as técnicas de detecção profunda de anomalias semi supervisionadas são bastante utilizadas. Tais técnicas impulsionam rótulos existentes de uma classe positiva visando separar os *outliers* (CHALAPATHY; CHAWLA, 2019).

Os *autoencoders* que são chamados codificadores automáticos demonstram dados em várias camadas ocultas, reestruturando os dados de entrada. Quando treinados apenas em instâncias de dados positivos, falham na hora de reconstruir as amostras de dados anômalos. As amostras de dados que produzem erros atípicos são consideradas *outliers*. Uma maneira de usar *autoencoders* é treiná-los de forma **semi supervisionada** com amostras de dados positivas, e com as amostras de treinamento equilibradas (CHALAPATHY; CHAWLA, 2019).

- As técnicas de **detecção profunda de anomalias não supervisionadas** detectam *outliers* exclusivamente com base nas propriedades específicas das instâncias de dados (CHALAPATHY; CHAWLA, 2019).

A grande vantagem em usar técnicas de aprendizagem profunda é sua capacidade específica em aprender características complexas diretamente dos dados, intratáveis por humanos manualmente. Isso permite definir um modelo que não dependa diretamente dessas características extraídas manualmente das imagens, que em algumas técnicas de aprendizado de máquina geralmente são necessárias (ANWAR et al., 2018).

Por conseguinte, nas próximas seções serão exploradas técnicas de aprendizado profundo que automatizam e otimizam o processo de extração de informações, além de permitir o aprendizado de representações mais complexas.

## 2.2 *Deep Learning*

O *aprendizado profundo* é um procedimento da visão computacional que seu funcionamento é baseado conforme a funcionalidade do cérebro humano que são baseados nas redes neurais ou estatísticas (GOODFELLOW; BENGIO; COURVILLE, 2016). Essa capacidade de aprendizagem automática de atributos exclui a necessidade de extrair características manuais das imagens médicas por especialistas no domínio (CHALAPATHY; CHAWLA, 2019).

O aprendizado profundo é um exemplo do modelo de aprendizado de máquina, onde o aprendizado profundo aperfeiçoa as representações aprendidas dos dados pressupostos, visando atingir ao máximo a separabilidade entre as classes. Por esse motivo que toda rede neural de aprendizagem profunda inicia com a mesma conjectura de inicialização aleatória e a cada iteração, os dados são propagados pela rede para calcular sua respectiva saída. Esta saída é comparada à saída desejada (por exemplo, determinar se um *pixel* em questão pertence a uma imagem ou não), e um erro é calculado para cada parâmetro da rede, para poder ser ajustado para uma melhor separação das amostras de treinamento na classe correta (JANOWCZYK; MADABHUSHI, 2016).

A seguir, é discutido a metodologia necessária no processamento dos dados usados ao aplicar aprendizado profundo a uma tarefa de análise de imagem de DP, o qual é baseado em (JANOWCZYK; MADABHUSHI, 2016).

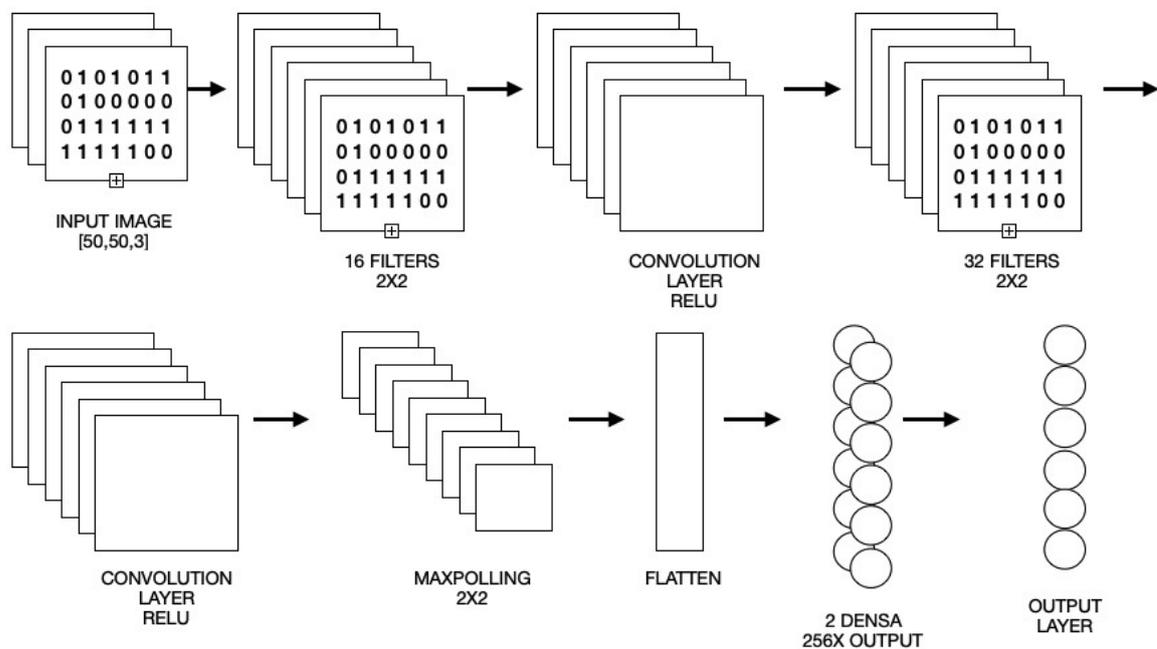
- **Casting:** É o momento de tomar várias decisões, como escolher o tamanho do *patch* de entrada, o número de camadas e atributos convolucionais, para projetar uma rede apropriada. Usar uma arquitetura existente nesse momento é crucial, pelo fato de ser um desafio difícil encontrar a configuração de rede mais bem-sucedida para um determinado problema e a quantidade de tempo concomitante para treinamento e teste da rede. Com isso, pode-se medir o desempenho de outras configurações de rede disponíveis em relação a um *benchmark* conhecido.
- **Geração de patch:** Uma vez que a rede neural é definida, os *patches*, que são os pedaços de cada imagem que foi dividida, precisam ser gerados para construir os conjuntos de treinamento e validação. Uma vez que a rede escolhida tem capacidade de discriminação limitada, a seleção de *patches* de imagem apropriados pode impactar no desempenho do modelo. Garantir que um conjunto de exemplares seja extraído das imagens é talvez um dos aspectos mais importantes para alavancar e utilizar com eficácia uma abordagem.
- **Treinamento:** O processo de treinamento vem logo depois que os dados estão extraídos e a rede neural está bem definida. Primeiro a série de *patches* de imagens são introduzidas na rede ao longo de uma série de épocas, com isso, uma derivada de erro é calculada e propagada de volta através da rede atualizando os pesos da rede. A taxa de aprendizagem é analisada temporalmente para que um mínimo local seja alcançado. Os pesos aprendidos que resultam no modelo, são armazenados para serem utilizados posteriormente no momento do teste.
- **Teste:** Ao enviar os *patches* de imagem para a rede, do mesmo tamanho usado durante o treinamento, obtêm-se uma previsão de classe a partir dos pesos aprendidos.

### 2.3 Redes Neurais Convolucionais

As *Redes Neurais Convolucionais* (CNNs) são particularidades de inspiração biológica de perceptrons de multicamadas, que também são baseadas na funcionalidade do cérebro humano. As CNNs buscam reconhecer padrões visuais, diretamente dos pixels da imagem, em alguns casos, é necessário um pré-processamento antes de fornecer as reproduções CNNs. Essas CNNs miram em pequenos segmentos das imagens de entrada, intitulados campos receptivos, utilizando neurônios de múltiplas camadas e pesos compartilhados em cada camada convolucional. Essas redes neurais profundas apontam para os campos receptivos da imagem de entrada, usam múltiplas camadas de neurônios com pesos compartilhados, visando extrair o máximo de características da imagem (ANWAR et al., 2018).

A estrutura da CNN é formada de convoluções, *pooling* máximo, camada densa e conectadas. Toda camada convolucional gera um esquema de características com dimensões diferentes e camadas ligadas reduzem o tamanho dos esquemas de características a serem transferidos para as camadas seguintes. As camadas completamente conectadas na saída produzem a previsão de classe necessária. O número de parâmetros necessários para definir uma rede depende da quantidade de camadas, filtros convolucionais e as suas dimensões, e a conexão entre eles. A fase de treinamento da rede garante que os melhores pesos possíveis sejam aprendidos (ANWAR et al., 2018). Na Figura 2, é ilustrado uma arquitetura CNN.

Figura 2 – Uma arquitetura de rede neural convolucional típica para classificação de imagens médicas



Fonte: Autor

O processo de convolução é uma estratégia para representar imagens maiores, desde um conjunto de características aprendidos funcionando como detectores de características. Para fazer isso, aplica-se uma camada de convolução 2D dos mapas de características de entrada (isto é, canais de imagem para a primeira camada), com um *kernel* de convolução quadrado (isto é, filtros ou elementos), é definido como

$$y_j = \tanh(\sum_i k_{ij} * x_i)$$

, onde  $x_i$  corresponde à  $i$ -ésima entrada do mapa de feições  $i$ ,  $k_{ij}$  é o *kernel* de convolução e  $y_j$  corresponde ao  $j$ -ésimo mapa de feições de saída. A função  $\tanh(\cdot)$  é usada para retificar propriedades que não dependem de covariâncias, como dispersão para acentuar as diferenças

entre os recursos de entrada e acelerar o aprendizado baseado em gradiente não linear na saída de convolução (CRUZ-ROA et al., 2014).

A camada de *pooling* permite reduzir a grande dimensão original da representação da imagem através de uma estratégia de subamostragem que suporta invariâncias de espaço local (CRUZ-ROA et al., 2014).

A camada conectada é frequentemente aplicada na camada superior de uma arquitetura CNN para capturar relacionamentos complexos entre filtros de alto nível. Neste estágio, as informações espaciais são ignoradas para aprender a correlação entre diferentes locais. Assim, a saída da camada de *pooling* é a entrada de uma camada conectada que as mistura em um esquema de características (CRUZ-ROA et al., 2014).

A etapa final é a camada de classificação que é uma camada conectada com um neurônio para cada uma das duas classes (benigna ou maligna) ativado por um modelo classificador *softmax*. Todo o modelo CNN é treinado utilizando a descida do gradiente estocástico para minimizar a função perda de função seguinte:

$$L(x) = -\log \left[ \frac{e^{x_i}}{\sum_j e^{x_j}} \right]$$

, onde  $x_i$  corresponde a saídas de camada integral, ligada multiplicado por parâmetros do modelo classificador. Assim, as saídas da CNN são as probabilidades logarítmicas de associação de classe (CRUZ-ROA et al., 2014).

Outro ponto importante é que o uso de compartilhamento de pesos torna possível a implementação da rede convolutiva de forma paralela. Um perceptron de múltiplas camadas pode aprender um mapeamento não-linear complexo, de alta dimensionalidade, restringindo seu projeto através da incorporação de conhecimento prévio sobre a tarefa considerada. Segundo, os pesos sinápticos e níveis de *bias* podem ser aprendidos circulando-se o algoritmo de *backpropagation* através do conjunto de treinamento (HAYKIN, 2007).

### 3 TRABALHOS RELACIONADOS

Atualmente, pesquisadores tem estudado e aplicado inúmeras técnicas de aprendizado profundo para desenvolver soluções que melhorem o processo de análise de imagens médicas. A seguir, serão investigados alguns desses trabalhos existentes na literatura.

Em (JANOWCZYK; MADABHUSHI, 2016), os autores utilizam uma aplicação de aprendizado profundo baseada na arquitetura *AlexNet* para a tarefa de análise de imagens através de sete métodos exclusivos usados como casos de uso: (a) segmentação de núcleos, (b) segmentação do epitélio, (c) segmentação dos túbulos, (d) detecção de linfócitos, (e) detecção de mitose, (f) detecção de carcinoma ductal invasivo e (g) classificação de linfoma. Os autores fazem uma avaliação em cada caso de uso, para criar um tutorial para a análise de imagens médicas na Patologia Digital.

Um dos *datasets* que os autores utilizaram foi o *Kaggle Breast Histopathology Images*<sup>1</sup>, o mesmo que está sendo utilizado no presente trabalho. Os autores reduziram a amostragem de suas imagens originais de  $40 \times 40$  por um fator de 16: 1, para uma ampliação aparente de 2,5x. Os autores testaram três abordagens diferentes para utilizar esses *patches*  $50 \times 50$  e inseri-los no domínio de solução de  $32 \times 32$ : o redimensionamento, onde todo o *patch*  $50 \times 50$  foi redimensionado para  $32 \times 32$ ; o corte, onde cada imagem  $50 \times 50$  foi cortada em um *sub-patch* de  $32 \times 32$  exatamente do centro para garantir que o rótulo da classe fosse retirada corretamente; e o recorte com rotações adicionais, para compensar o conjunto de treinamento altamente desequilibrado, onde a classe negativa é representada mais de 3 vezes. Os autores artificialmente sobrem amostraram o gênero positivo adicionando rotações adicionais. Como resultado, foram utilizadas rotações de 0, 45, 90, 135, 180 graus, com seus espelhos para o conjunto de treinamento para a classe positiva. Os autores utilizaram o *F-score* e a acurácia balanceada para comparar os resultados com (CRUZ-ROA et al., 2014). Os resultados obtidos foram respectivamente: *F-score* e acurácia balanceada (71,78% e 84,23%) do *dataset* original, e (76,48% e 84,68%) para a arquitetura *AlexNet* com corte das imagens. É possível ver o uso da rede fornece *F-score* superior e também há um equilíbrio de precisão maior. O autor considera que o redimensionamento das imagens forneceu os melhores resultados. (JANOWCZYK; MADABHUSHI, 2016).

Em (WANG et al., 2020), seis modelos de aprendizagem profunda foram investigados a partir da consistência do desempenho em quatro conjuntos de dados de mamografia, incluindo três *datasets* públicos: Banco de Dados Digital para Mamografia de Triagem (DDSM), *IN breast*, *Mammographic Image Analysis Society* (MIAS) (Sociedade de Análise de Imagens Mamográficas) e um conjunto de dados privado (UKy). As mamografias contêm calcificações, massas ou ambos foram rotulados como benignos ou malignos. Dois desses modelos são *end2end*, um é *frCNN\_cad*, e os três modelos criados pelos autores do artigo. A área sob a curva característica de operação do receptor (auROC) foi usada para avaliar o desempenho dos modelos. Todos os modelos foram treinados no maior banco de dados de mamografia disponível ao público DDSM, foram avaliados e os resultados foram registrados quanto a precisão dos modelos em DDSM. Finalmente, cada modelo foi testado em três diferentes conjuntos de dados e os resultados foram comparados com os resultados da validação para avaliar a consistência do desempenho. O *end2end* e os três novos modelos projetados usaram a técnica de *transfer learning*, que pré-treina os modelos CNN com imagens naturais do ImageNet, a qual alimenta as mamografias como extratores de características, assim, classificando os mapas de características usando um classificador CNN. Por fim, transfere os modelos para outro domínio de imagens posteriormente. Esta técnica foi comprovada para aliviar o problema do pequeno tamanho dos dados. Os modelos utilizados como extratores de recursos foram o AlexNet, VGG16 e ResNet50. O modelo *frCNN\_cad* usou o método de aprendizado baseado em instância, já os modelos *end2end* usavam VGG16 e ResNet50 como *backbone*.

<sup>1</sup> <https://www.kaggle.com/paultimothymooney/breast-histopathology-images>

Os três modelos criados pelos autores em (WANG et al., 2020) obtiveram um resultado satisfatório entre 0,71 e 0,79 de (auROC) no mesmo conjunto de dados de validação, já os três modelos publicados obtiveram um resultado melhor, entre 0,88 e 0,95 no conjunto de dados de validação. No entanto, os mesmos critérios de avaliação de todos os seis modelos nos três conjuntos de dados de teste externos foram significativamente reduzidos apenas entre 0,44 e 0,65, mostrando haver uma inconsistência de desempenho nos resultados entre os conjuntos de dados e modelos, e que nem sempre o alto desempenho de um determinado modelo de aprendizagem profunda pode ser transferido ou generalizado para outro conjunto de dados externo, sem antes ter sido avaliado e validado antes de ser aplicado na prática clínica. Neste estudo, o conjunto de dados DDSM é bastante diferente dos dados de teste em muitos aspectos. Por exemplo, o DDSM é um conjunto de dados de mamografia de tela digital, mas *INbreast* e *UKy* são conjuntos de dados de mamografia digital, com resposta de densidade óptica diferente à exposição de raios-x. Portanto, as escolhas dos parâmetros de aquisição da imagem também podem afetar as características da imagem. Os resultados desse estudo demonstram que modelos de aprendizagem treinados em um conjunto de dados limitado não funcionam bem em conjuntos de dados com diferentes distribuições na população de pacientes, características da doença e sistemas de imagem. Mas também, o estudo mostra que o treinamento com dados mistos pode melhorar o desempenho do modelo.

Em (NAWAZ et al., 2018), é investigado uma abordagem baseada em *transfer learning* para a classificação de imagens histológicas coradas com hematoxilina e eosina (H & E) de câncer de mama. A rede neural convolucional profunda *ALEXNET* é ajustada para imagens de câncer de mama. O conjunto de dados de câncer de mama fornecido consiste em imagens de alta qualidade ( $2048 \times 1536$  píxeis) não comprimidas e marcadas com hematoxilina e eosina (H & E), uma técnica usada para a coloração de imagens histológicas. Esses dados compõem-se em 400 imagens (100 para cada classe) e cada imagem é atribuída a um dos quatro rótulos de classe: benigno, normal, InSitu e invasivo. As 400 imagens foram particionadas em dois conjuntos, X1: consistindo em 320 conhecidas como conjunto de treinamento e X2: consistindo em 80 designadas como conjunto de validação para a avaliação. Destaca-se que a normalização da mancha tende a dar uma melhora na análise quantitativa. É observado também que essa normalização resultou em ganho de desempenho quando comparado ao treinamento em imagens não normalizadas. As imagens de treinamento e validação são separadas aleatoriamente em conjuntos separados. A versão ampliada do modelo treinado pelos autores foi chamada *TK-ALEXNET*.

As camadas de entrada de *ALEXNET* foram alteradas para o tamanho  $512 \times 512$  em vez de usar  $227 \times 227$ , e outras configurações na rede como remover todas as camadas, adicionar convolução, ativação e camada máxima de *pooling* foram feitas visando evitar *overfitting*. A avaliação de desempenho foi conduzida usando precisão e sensibilidade, e os resultados apresentados são para o conjunto de validação e o conjunto de teste oculto do Grande Desafio ICIAR-2018 (Análise e reconhecimento de imagens de conferências internacionais) em imagens de histologia do câncer de mama. Os autores tiveram acesso apenas à precisão do conjunto

oculto, por isso não conseguiram apresentar métricas de desempenho para as classes individuais. A precisão do modelo *TK-ALEXNET* é de 75,73% para quatro classes, já a precisão da imagem do *TK-ALEXNET* para quatro classes é de 81,25% no conjunto de dados de validação. Indicativos empíricos mostram que esse procedimento fornece uma precisão geral de 57% no conjunto de teste oculto e essa diminuição na precisão pode ter resultado devido à complexidade do conjunto de dados. Como trabalhos futuros, os autores pretendem expandir a pesquisa visando melhorar o desempenho usando arquiteturas profundas personalizadas para classificação de câncer.

Ao analisar os trabalhos relacionados, nota-se a importância do desenvolvimento da pesquisa proposta por esse trabalho. Mesmo com a grande relevância dos resultados obtidos com esses modelos e conjuntos de dados utilizados, é necessária uma pesquisa e avaliação com modelos e hiper-parâmetros diferentes, pelo fato de que, é citado pela literatura e mostrado nos trabalhos relacionados a importância de criar protótipos para outros casos específicos na área de análise de imagens médicas.

## 4 METODOLOGIA

Nesta seção, é apresentada a metodologia deste trabalho, que consiste na aplicação de técnicas de aprendizado profundo na detecção de anomalias em imagens médicas utilizando redes neurais convolucionais e classificação binária. A seguir, são apresentados as etapas desenvolvidas no decorrer da pesquisa.

### 4.1 Preparação dos Dados

Inicialmente, será apresentado o *dataset Kaggle Breast Histopathology Images*<sup>2</sup>, que foi utilizado para treinamento e avaliação de desempenho da abordagem proposta para classificação binária supervisionada de imagens médicas. Este *dataset* é composto de exemplos positivos e negativos para o *Carcinoma Ductal Invasivo* (IDC) o qual é o subtipo mais comum de todos os cânceres de mama. Para atribuir um grau de agressividade a uma amostra de imagem inteira, os patologistas geralmente se concentram nas regiões que contêm a anomalia. Como resultado, uma das etapas comuns desse pré-processamento para classificação automática de agressividade é delinear as regiões exatas do IDC dentro de uma imagem completa.

O conjunto de dados original utilizado em (CRUZ-ROA et al., 2014) consistia em 162 imagens de *slides* inteiros de amostras de câncer de mama escaneados a 40x. Para um maior aproveitamento do conjunto de dados, as imagens obtidas foram extraídas em *patches*, sendo eles: 277.524 *patches* de tamanho 50 x 50 (198.738 IDC benignos e 78.786 IDC malignos).

Exemplos de tecidos classificados como IDC malignos e benignos, que serão usados nos conjuntos de treinamento e teste são mostrados na Figura 3.

No pré-processamento dos *patches*, cada *patch* de imagem é carregado como uma imagem RGB e normalizado para uma média de zero e variância de um. Esta etapa remove as correlações

<sup>2</sup> <https://www.kaggle.com/paultimothymooney/breast-histopathology-images>

Fonte: (JANOWCZYK; MADABHUSHI, 2016)

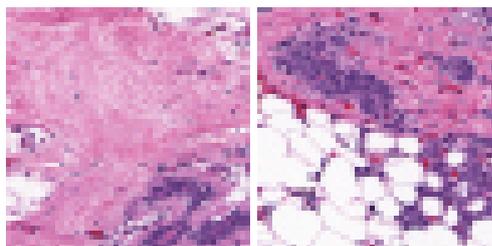


Figura 3 – Imagens microscópicas de tecidos mamários, onde a imagem da esquerda representa tecidos não carcinoma (benigno) e a imagem da direita representa tecidos com carcinoma (maligno)

de píxeis brutos, permitindo focar em propriedades que não dependem de covariâncias, como dispersão. A motivação para fazer isso é acentuar as diferenças entre os recursos de entrada e acelerar o aprendizado baseado em gradiente (CRUZ-ROA et al., 2014).

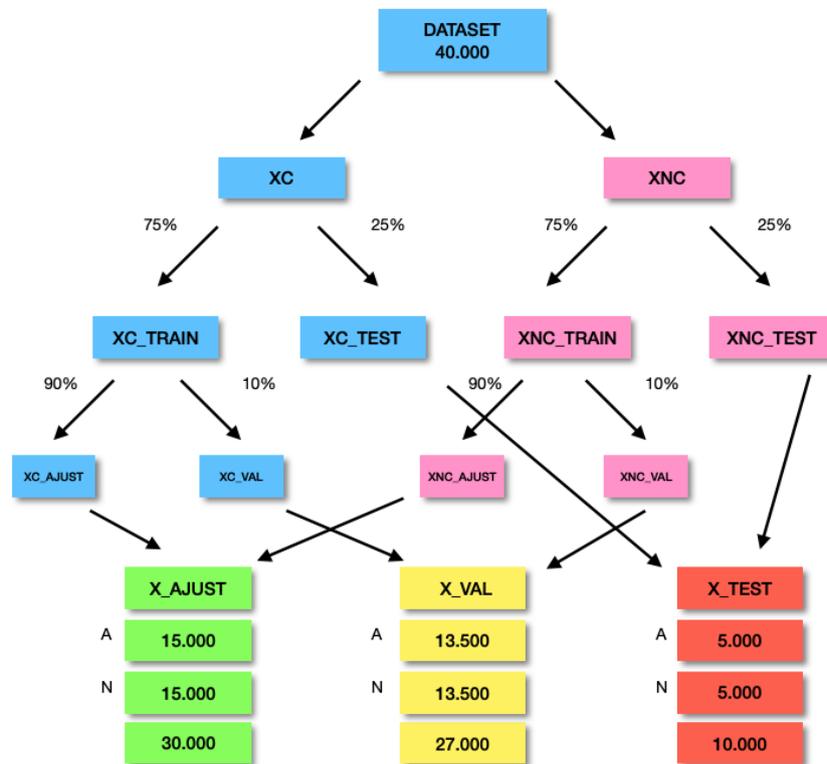
O conjunto de dados foi dividido em conjuntos de treinamento, validação e teste, onde do *dataset* de entrada foram obtidos 40.000 mil *patches* de imagens, devido à limitação de memória do *Kaggle* foram utilizados apenas 40.000 mil dos 277.524 mil *patches*. Esse valor foi dividido em dois conjuntos, o XC (20.000 *patches*) e XNC (20.000 *patches*), e depois transformado em dois *arrays*, onde o XC é identificado como maligno por conter anomalia e o XNC como benigno por não conter nenhuma anomalia. Desses 20.000 exemplos de cada conjunto, cada classe é dividida em quatro *datasets*: o *dataset* de treinamento (XC\_train — 75%) com 15.000 *patches* de tamanho 50x50, o *dataset* de teste (XC\_test — 25%) com 5.000, o *dataset* de ajuste (XC\_ajust — 90%) formado por 13.500 *patches* a partir do *dataset* de treinamento que contém 15.000, e o *dataset* de validação (XC\_validation — 10%) formado com 1.500 imagens do *dataset* de ensaio. É feito o mesmo processo para o XNC, dividimos o XNC em dois *datasets*: o de treinamento XNC\_train (75%) e o de Teste XNC\_test (25%), o XNC\_train é dividido em XNC\_ajust (90%) e XNC\_validation (10%). Em seguida, todos os conjuntos de treinamento, teste, validação e avaliação são concatenados em: X\_ajust, X\_test e X\_val, para que seguidamente possam ser treinados, ajustados e avaliados. As etapas de divisão do conjunto de dados serão mostradas a seguir na Figura 4.

O conjunto de dados foi dividido dessa forma, para manter a separabilidade das classes, visando manter os dados balanceados. Logo depois, os modelos foram treinados com o X\_train, ajustados com o X\_ajust, testados com o X\_test e avaliados com o X\_val. A seguir, é mostrado como o experimento foi avaliado.

## 4.2 Modelagem e Avaliação

Esta seção descreve os processos envolvidos para modelagem e avaliação de desempenho aplicada à abordagem proposta para a detecção de anomalias em imagens médicas. Antes de falar sobre a escolha da arquitetura da aplicação é necessário abordar as métricas que serão utilizadas.

Figura 4 – (Na figura mostra como foi dividido o conjunto de dados e depois concatenado. O 'A' equivale às imagens que são anômalas e o 'N' as que não, e no final é mostrado o valor da concatenação das duas.)



Fonte: Autor

As etapas aplicadas nesta pesquisa são descritas brevemente nas subseções seguintes.

#### 4.2.1 Métricas de avaliação

Um sistema típico de análise de imagens médicas é avaliado usando diferentes medidas-chave de desempenho, como exatidão, pontuação F1, precisão, *recall*, sensibilidade, especificidade e coeficiente de dados (JANOWCZYK; MADABHUSHI, 2016).

Na rede criada será calculada a matriz de confusão para derivar a precisão, sensibilidade e especificidade, visando entender o desempenho do modelo em um nível mais profundo, assim como avaliado para as arquiteturas CNNs presentes em (JANOWCZYK; MADABHUSHI, 2016).

De acordo com (JANOWCZYK; MADABHUSHI, 2016) matematicamente, essas medidas são calculadas como:

- **Verdadeiro Positivo (VP):** Representam o número de instâncias pertencentes a uma determinada classe e que foram classificados como pertencentes dessa classe.

- **Falso Positivo (FP):** Representam o número de instâncias não pertencentes a uma determinada classe e que foram classificados como pertencentes dessa classe.
- **Verdadeiro Negativo (VN):** Representam o número de instâncias que não pertencem a uma determinada classe e que foram classificados como não pertencentes dessa classe.
- **Falso Negativo (FN):** Representam o número de instâncias que pertencem a uma determinada classe e que foram classificados como não pertencentes dessa classe.

(JANOWCZYK; MADABHUSHI, 2016) citam as principais métricas de avaliação para sistema de análise de imagens médicas. A seguir será feita uma breve explicação das métricas mais comuns, as mesmas que serão adotadas na pesquisa.

- **Acurácia:** É a taxa de acerto do classificador, pode ser representada pela seguinte equação:

$$ACC = \frac{FP + FN}{VP + FP + FN + FN} \quad (1)$$

- **Precisão:** É a dimensão de instâncias positivas classificadas corretamente entre os positivos, pode ser representada pela equação seguinte :

$$PREC = \frac{VP}{VP + FP} \quad (2)$$

- **Recall:** É a taxa de acerto nas instâncias da classe positiva, pode ser representada pela seguinte equação:

$$REC = \frac{VP}{VP + FN} \quad (3)$$

- **FPR:** É a taxa de avaliação da capacidade do método de detectar resultados negativos e rejeitá-los, pode ser representada pela seguinte equação:

$$FPR = \frac{FP}{FP + VN} \quad (4)$$

- **F-Score:** É uma média calculada com base na precisão e na sensibilidade, pode ser representada pela seguinte equação:

$$F1 = 2 * \frac{PREC * REC}{PREC + REC} \quad (5)$$

- **BAC:** É uma média calculada de todos os acertos divididos por todos os acertos mais os erros (CRUZ-ROA et al., 2014). A métrica pode ser representada pela seguinte equação:

$$BAC = \left( \frac{REC + SENS}{2} \right) \quad (6)$$

Após o treinamento das técnicas de aprendizado profundo descritas na seção 2.2, a detecção de anomalias com convolução, os modelos foram submetidos ao processo de avaliação.

#### 4.2.2 Seleção de Hiper-parâmetros

Para avaliar os *patches* de imagens, utilizamos Redes Neurais Convolucionais, que usa a aplicação de detectores de características locais ou filtros em toda a imagem para medir a correspondência entre *patches* de imagens individuais e padrões de assinatura no conjunto de treinamento (CRUZ-ROA et al., 2014). Então, a etapa de *pooling* é aplicada para reduzir a dimensionalidade do espaço de características. Os *patches* de imagens do *Kaggle Breast Histopathology Images* são usados como entradas para uma arquitetura CNN, onde algumas camadas são usadas para convolução e outras para o *pooling* enquanto a camada restante está totalmente conectada. Para a avaliação das técnicas de ML e DL, foram utilizadas a biblioteca Keras<sup>3</sup>, que fornece uma *interface* para as redes neurais, o Tensorflow<sup>4</sup> e outras bibliotecas. Para definir a arquitetura final para o modelo de classificação, 18 arquiteturas foram avaliadas variando-se os parâmetros das camadas convolucionais.

Inicialmente, fixou-se a camada densa com 256 neurônios e variou-se o número de convoluções entre 1 e 2 camadas. O número de filtros convolucionais entre 16, 32 e 64 para 1 camada de convolução e para 2 números fixou o filtro convolucional para 16x32. A dimensão dos filtros convolucionais também foi variada entre 2x2, 3x3 e 4x4. No outro bloco de experimentos, adicionamos o *pooling* após cada camada convolucional. No último bloco de experimentos, variaram-se os parâmetros da camada densa entre 1 e 2, e o número de neurônios por camada densa entre 128, 256 e 512, nesse caso, os parâmetros fixados foram os que maximizaram o resultado: a quantidade de convoluções em 2, o número de filtros convolucionais para 16x32 com dimensão 2x2. Na arquitetura utiliza-se 256 filtros de convolução fixados na camada densa com a função de ativação *relu*, e na camada totalmente conectada utiliza a função *sigmoid* para extrair os resultados da rede neural. Durante o treinamento a função de custo *binary cross entropy*

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)) \quad (7)$$

<sup>3</sup> <https://keras.io/>

<sup>4</sup> <https://www.tensorflow.org/>

Tabela 1 – Avaliação de desempenho das arquiteturas de Redes Neurais Convolucionais

Arq.	Num Filtros	Dimensão	Pooling	Camadas Densas	Neurônios	Média (ACC)
1	16	2 x 2	Não	1	256	0,8016
2	32	3 x 3	Não	1	256	0,7911
3	64	4 x 4	Não	1	256	0,7864
4	[16, 32]	2 x 2	Não	1	256	0,801
5	[16, 32]	3 x 3	Não	1	256	0,8015
6	[16, 32]	4 x 4	Não	1	256	0,7985
7	16	2 x 2	Sim	1	256	0,8049
8	32	3 x 3	Sim	1	256	0,7895
9	64	4 x 4	Sim	1	256	0,7954
10	[16, 32]	2 x 2	Sim	1	256	0,8105
11	[16, 32]	3 x 3	Sim	1	256	0,8021
12	[16, 32]	4 x 4	Sim	1	256	0,7998
13	[16, 32]	2 x 2	Sim	1	128	0,809
14	[16, 32]	2 x 2	Sim	1	256	0,8094
15	[16, 32]	2 x 2	Sim	1	512	0,8057
16	[16, 32]	2 x 2	Sim	2	128	0,815
<b>17</b>	<b>[16, 32]</b>	<b>2 x 2</b>	<b>Sim</b>	<b>2</b>	<b>256</b>	<b>0,8233</b>
18	[16, 32]	2 x 2	Sim	2	512	0,8074

onde  $y_i \in \{0, 1\}$  e  $p(y_i)$  é a probabilidade de ocorrência do padrão  $i$ , e a função de otimização *adam* são utilizadas. Nos experimentos foi adicionado o *EarlyStopping*<sup>5</sup>, um método de regularização do Keras, no processo de treinamento, que tem a funcionalidade de parar o treinamento quando uma métrica monitorada parar de aprimorar. Na Tabela 1 são detalhadas as arquiteturas avaliadas durante os experimentos.

Todas as arquiteturas foram executadas 3 vezes para 100 épocas. Visando obter a melhor acurácia entre as arquiteturas, após cada execução do modelo, foi calculado a média da acurácia para às três execuções. Após o processo de avaliação da arquitetura de rede neural convolucional, a arquitetura final é definida (17). A arquitetura final apresentada na Tabela 2 obteve o melhor desempenho para a métrica acurácia (ACC) durante o experimento. Esta arquitetura de rede foi definida com base no desempenho e contém um total de 1.059.825 parâmetros treináveis. No modelo é inicialmente aplicado um processo de convolução com 16 filtros na imagem de entrada. Depois disso, é aplicado outro processo de convolução com 32 filtros. Em cada camada seguinte de convolução, é aplicado a operação de *MaxPooling* com filtro 2x2, que reduzirá o tamanho da imagem. Por fim, esses valores são unidos em um *array* pela operação *Flatten* para serem utilizados como a entrada das redes neurais tradicionais densas na camada de saída.

Para permitir que os pesquisadores avaliem os seus métodos e possa fazer uma comparação, na prática, este trabalho está disponível no link<sup>6</sup>. Na próxima seção é apresentado os resultados obtidos durante a avaliação da rede neural convolucional.

<sup>5</sup> [https://keras.io/api/callbacks/early\\_stopping/](https://keras.io/api/callbacks/early_stopping/)

<sup>6</sup> <https://github.com/robsonsants/CNN-TCC2>

Tabela 2 – Arquitetura de rede neural convolucional que obteve o melhor desempenho durante a avaliação.

<b>Camada</b>	<b>Tipo</b>	<b>Forma de saída</b>	<b>Parâmetros</b>
<i>Camada de entrada</i>	(Conv2D)	(49, 49, 16)	208
<i>Camada Oculta 1</i>	(MaxPooling2D)	(24, 24, 16)	0
<i>Camada Oculta 2</i>	(Conv2D)	(23, 23, 32)	2080
<i>Camada Oculta 3</i>	(MaxPooling2D)	(11, 11, 32)	0
<i>Camada Oculta 4</i>	(Flatten)	(3872)	0
<i>Camada Oculta 5</i>	(Dense)	(256)	991488
<i>Camada Oculta 6</i>	(Dense)	(256)	65792
<i>Camada de Saída</i>	(Dense)	(1)	257

## 5 RESULTADOS

Nesta seção é apresentado os principais resultados obtidos a partir da avaliação de desempenho da abordagem proposta para detecção de imagens de câncer de mama do tipo IDC para auxiliar no processo de diagnóstico apoiando os profissionais de saúde na tomada de decisões. Após a escolha da arquitetura final, o experimento foi executado novamente com as configurações apresentadas na Tabela 2, visando obter mais resultados com a arquitetura de melhor desempenho. As medidas de desempenho apresentadas são: acurácia (ACC), curva (AUC), *False Positive Rate* (FPR), precisão (PREC), Recall (REC), *F-score* (F1) e acurácia balanceada (BAC). Na Tabela 3, é apresentada a média dos resultados do classificador CNN para duas execuções, a primeira utilizando o *EarlyStopping* e a segunda sem o *EarlyStopping*. É apresentado também o desvio padrão de cada métrica para a arquitetura 17, executados 30 vezes com 2 filtros convolucionais de [16,32] com dimensão 2x2, 2 camadas densas com 256 neurônios cada uma e o *pooling* após cada camada convolucional.

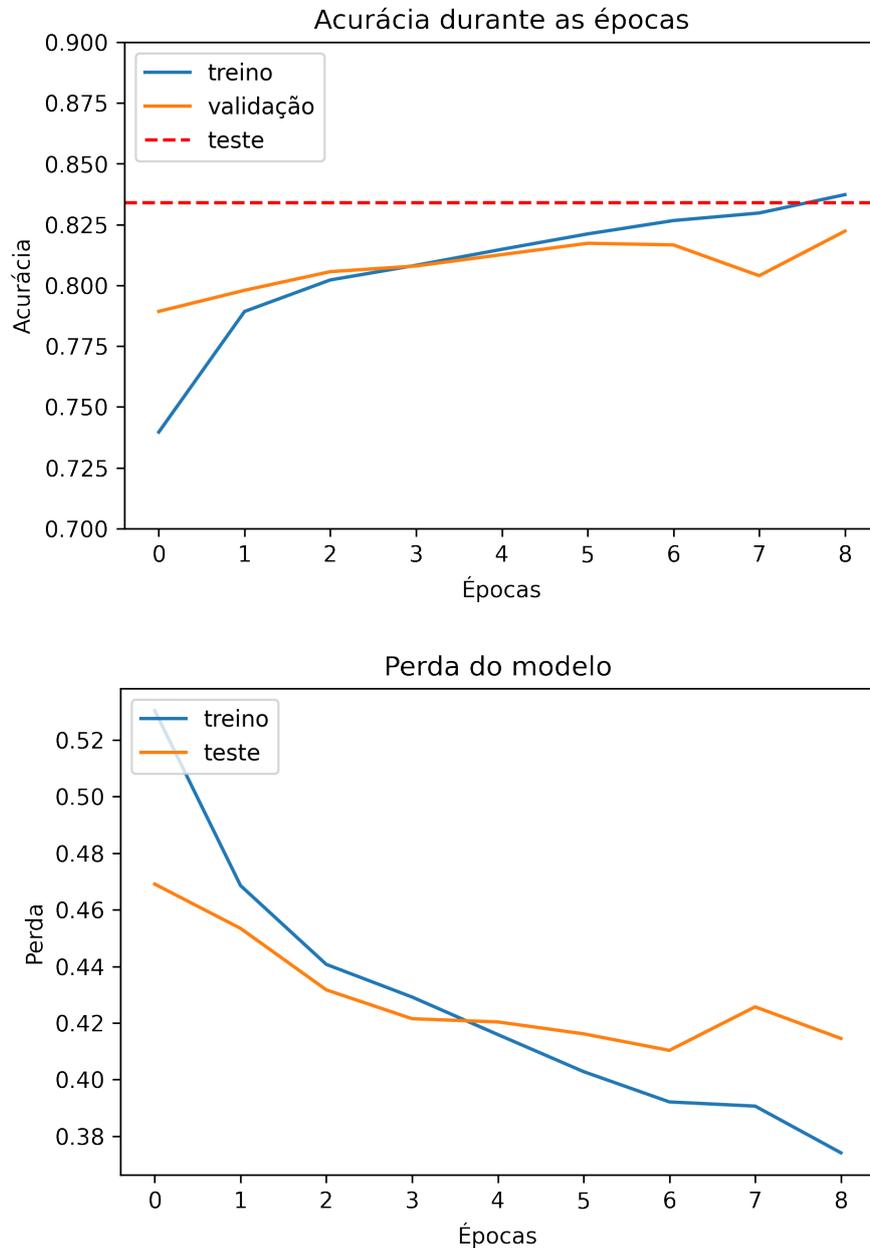
Tabela 3 – Resultados obtidos para a classificação de IDC.

COM EARLY STOPPING	ACC	AUC	FPR	PREC	REC	F1
Média	0,8138	0,8994	0,1690	0,8322	0,8089	0,8168
Desvio Padrão	0,0122	0,0063	0,0387	0,0623	0,0488	0,0122
SEM EARLY STOPPING	ACC	AUC	FPR	PREC	REC	F1
Média	0,7947	0,8611	0,2069	0,7919	0,7965	0,7941
Desvio Padrão	0,0049	0,0044	0,0052	0,0062	0,0063	0,0049

As arquiteturas construídas alcançaram resultados aproximados para as métricas de desempenho consideradas. Em alguns testes com diferentes camadas ocorre o *overfitting* dos resultados quando não é utilizado o *EarlyStopping* no experimento. Na Tabela 3 é apresentado a média dos resultados e o desvio padrão que é extremamente baixo, demonstrando que os valores dos modelos obtidos nos resultados tendem a estar próximos da média do conjunto. Na Figura 5 é mostrado o melhor desempenho da acurácia de uma das 30 execuções e a função de perda durante o treinamento com o uso do *EarlyStopping*. Já na Figura 6 é mostrado o melhor

desempenho de acurácia e a função de perda sem o uso do *EarlyStopping*. É possível notar o *overfitting* quando não é utilizado o *EarlyStopping*.

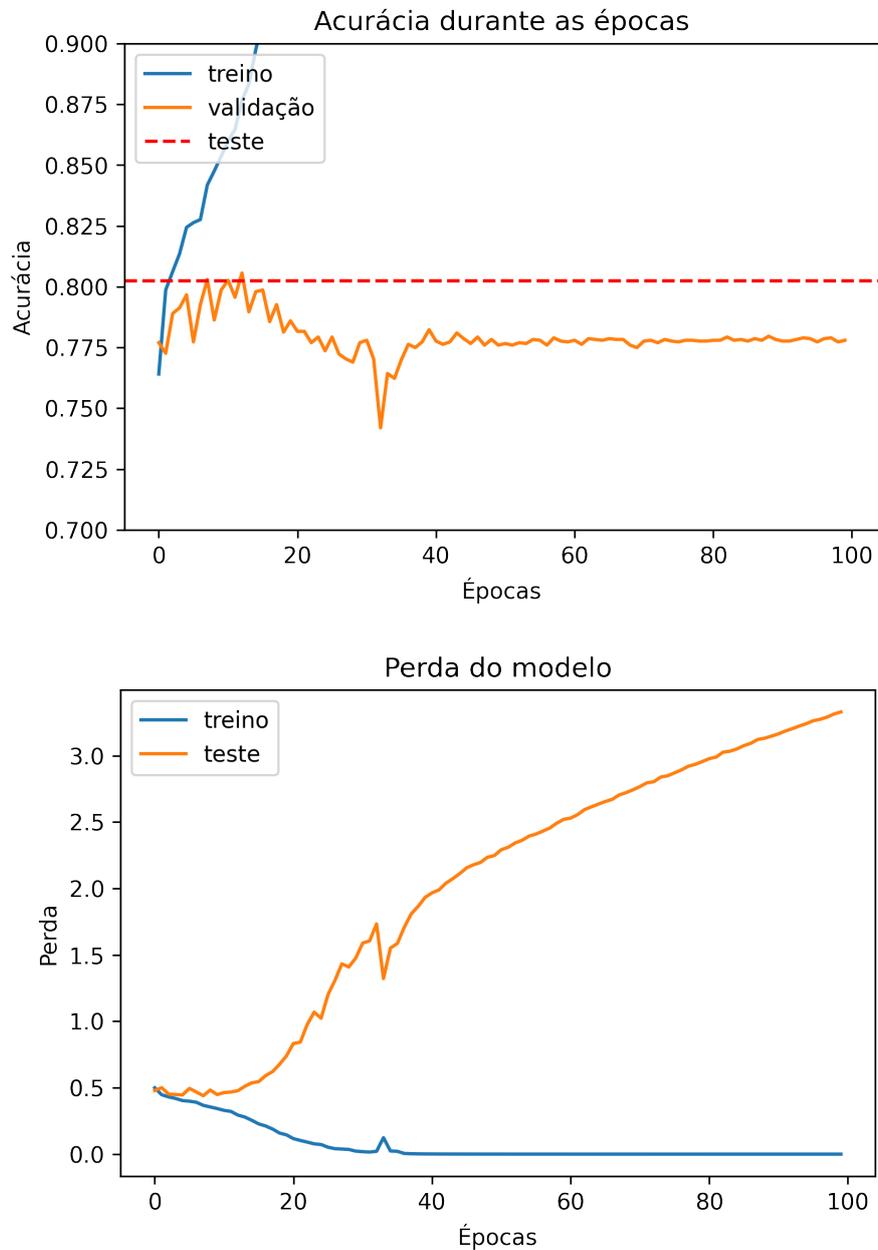
Figura 5 – (Acurácia e loss function durante o treinamento (Com *EarlyStopping*))



Fonte: Autor

O desempenho da arquitetura proposta foi ainda comparada com o trabalho (JANOWCZYK; MADABHUSHI, 2016) e (CRUZ-ROA et al., 2014), conforme a Tabela 4. Os resultados de (CRUZ-ROA et al., 2014) são avaliados para 10 arquiteturas, uma delas utiliza rede neural convolucional simples e as outras utilizam recursos manuais para avaliar a eficiência de sua abordagem, ou seja, há uma avaliação com predições feito com a ajuda de patologistas. O melhor desempenho foi da rede CNN, avaliado pela acurácia balanceada (BAC) e *F-score*.

Figura 6 – (Acurácia e loss function durante o treinamento (Sem EarlyStopping))



Fonte: Autor

Já os autores em (JANOWCZYK; MADABHUSHI, 2016), utilizam uma rede baseada na arquitetura da AlexNet e testam três abordagens diferentes com base no modelo AlexNet: o redimensionamento dos *patches* de imagens, o corte, e o recorte + rotações adicionais. O melhor desempenho dessas três abordagens foi a do redimensionamento dos *patches*. É possível notar que a abordagem proposta apresenta resultados comparáveis em BAC e supera os resultados em comparação ao *F-score*.

Tabela 4 – Comparação com a Literatura

<b>Abordagem</b>	<b>BAC (%)</b>	<b>F1 (%)</b>
(JANOWCZYK; MADABHUSHI, 2016)	84,68%	76,48%
(CRUZ-ROA et al., 2014)	84,23%	71,80%
<b>Abordagem proposta (10)</b>	82,24%	<b>81,58%</b>

## 6 CONCLUSÕES E TRABALHOS FUTUROS

O *aprendizado profundo* vem se tornando uma área importante com aplicações em áreas diversas, na análise de imagens médicas não é diferente. Considerando a quantidade de trabalhos que buscam melhorar o processo de detecção de anomalias por meio das CNNs, é relevante buscar soluções que lidem com a detecção de anomalias, tendo em vista que a maioria dos trabalhos presentes na literatura utilizam o aprendizado profundo para a tarefa de análise de imagens.

Neste trabalho foi apresentado e avaliado uma abordagem de redes neurais convolucionais para a classificação binária e detecção de anomalias de imagens histopatológicas de câncer de mama. O dataset *Kaggle Breast Histopathology Images*, foi utilizado para a modelagem, treinamento e avaliação da rede proposta. A abordagem proposta obteve resultados comparáveis de BAC e superiores em *F-score* para os dois trabalhos comparados. O modelo atingiu uma acurácia balanceada de 82,24% e *F-score* de 81,58%, durante as execuções do experimento.

Como trabalhos futuros, propõe-se em explorar outras técnicas de aprendizado profundo para a tarefa de análise de imagens médicas, envolvendo outros *datasets* com o intuito de comparar as técnicas utilizadas, visando melhorar a arquitetura proposta e resultados melhores. Além disso, planeja-se o desenvolvimento de um *software* utilizando a abordagem proposta para a detecção de anomalias em imagens de câncer de mama, para posteriormente ser aplicado em ambientes clínicos reais.

## REFERÊNCIAS

- ANWAR, S. M. et al. Medical image analysis using convolutional neural networks: a review. **Journal of medical systems**, Springer, v. 42, n. 11, p. 1–13, 2018.
- CHALAPATHY, R.; CHAWLA, S. Deep learning for anomaly detection: A survey. **arXiv preprint arXiv:1901.03407**, 2019.
- CRUZ-ROA, A. et al. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Medical Imaging 2014: Digital Pathology**. [S.l.], 2014. v. 9041, p. 904103.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>.
- HAYKIN, S. **Redes Neurais: Princípios e Prática**. Artmed, 2007. ISBN 9788577800865. Disponível em: <<https://books.google.com.br/books?id=bhMwDwAAQBAJ>>.

JANOWCZYK, A.; MADABHUSHI, A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. **Journal of pathology informatics**, Wolters Kluwer–Medknow Publications, v. 7, 2016.

NAWAZ, W. et al. Classification of breast cancer histology images using alexnet. In: SPRINGER. **International conference image analysis and recognition**. [S.l.], 2018. p. 869–876.

WANG, X. et al. Inconsistent performance of deep learning models on mammogram classification. **Journal of the American College of Radiology**, Elsevier, v. 17, n. 6, p. 796–803, 2020.

## 7 ANEXOS

Tabela 5 – Resultados para 30 execuções com Early Stopping

	ACC	AUC	FPR	PREC	REC	F1
EXEC 1	0,8175	0,9008	0,2111	0,768	0,8523	0,8079
EXEC 2	0,812	0,8972	0,1283	0,8922	0,7688	0,8259
EXEC 3	0,809	0,8954	0,236	0,7236	0,8726	0,7911
EXEC 4	0,8264	0,9019	0,1887	0,802	0,8431	0,822
EXEC 5	0,8309	0,9069	0,1508	0,857	0,8144	0,8352
EXEC 6	0,8341	0,9068	0,1753	0,8196	0,844	0,8316
EXEC 7	0,8035	0,8892	0,138	0,8842	0,7613	0,8181
EXEC 8	0,8071	0,9031	0,2493	0,6946	0,8962	0,7826
EXEC 9	0,8206	0,904	0,1917	0,8006	0,8339	0,8169
EXEC 10	0,8133	0,9018	0,1328	0,8866	0,7732	0,826
EXEC 11	0,8226	0,8982	0,1915	0,7996	0,8381	0,8184
EXEC 12	0,8152	0,9047	0,1258	0,894	0,7722	0,8286
EXEC 13	0,8207	0,9004	0,154	0,8572	0,7988	0,827
EXEC 14	0,7876	0,8918	0,1229	0,9062	0,7324	0,8101
EXEC 15	0,8008	0,8903	0,14794	0,8736	0,7625	0,8143
EXEC 16	0,7931	0,8858	0,1307	0,8962	0,7429	0,8124
EXEC 17	0,8165	0,8976	0,166	0,8426	0,8007	0,8211
EXEC 18	0,7934	0,8889	0,1471	0,8776	0,7511	0,8094
EXEC 19	0,8324	0,9077	0,1521	0,8546	0,8182	0,836
EXEC 20	0,8253	0,9018	0,1606	0,846	0,8123	0,8288
EXEC 21	0,8037	0,8946	0,1411	0,8806	0,7632	0,8177
EXEC 22	0,8223	0,9088	0,2156	0,7556	0,8719	0,8096
EXEC 23	0,8186	0,9036	0,1297	0,8884	0,7795	0,8304
EXEC 24	0,8287	0,9076	0,2092	0,7634	0,878	0,8167
EXEC 25	0,8173	0,9044	0,2196	0,7514	0,8654	0,8044
EXEC 26	0,8162	0,8989	0,2175	0,7564	0,8591	0,8045
EXEC 27	0,7955	0,898	0,1212	0,9054	0,7422	0,8157
EXEC 28	0,8169	0,9033	0,2249	0,7408	0,8737	0,8018
EXEC 29	0,8112	0,8896	0,1589	0,855	0,7861	0,8191
EXEC 30	0,8035	0,8981	0,1306	0,8926	0,7575	0,8195
Média (%)	81,39%	89,94%	16,90%	83,22%	80,89%	81,68%
DP (%)	1,22%	0,63%	3,87%	6,23%	4,88%	1,22%

Tabela 6 – Resultados para 30 execuções sem Early Stopping

	ACC	AUC	FPR	PREC	REC	F1
EXEC 1	0,8025	0,8664	0,1997	0,7988	0,8047	0,8017
EXEC 2	0,7992	0,8613	0,2042	0,7934	0,8027	0,798
EXEC 3	0,7976	0,8637	0,2062	0,791	0,8015	0,7962
EXEC 4	0,801	0,8714	0,1977	0,803	0,7998	0,8013
EXEC 5	0,8006	0,8665	0,2034	0,7938	0,8047	0,7992
EXEC 6	0,7846	0,8526	0,217	0,7814	0,7864	0,7839
EXEC 7	0,7922	0,8579	0,2101	0,7882	0,7945	0,7913
EXEC 8	0,8007	0,8595	0,2017	0,7966	0,8031	0,7998
EXEC 9	0,7898	0,856	0,2151	0,7812	0,7948	0,7879
EXEC 10	0,7956	0,8581	0,2041	0,796	0,7953	0,7956
EXEC 11	0,7979	0,8645	0,2002	0,801	0,796	0,7985
EXEC 12	0,7971	0,8654	0,2044	0,7944	0,7987	0,7965
EXEC 13	0,7922	0,8588	0,212	0,7848	0,7965	0,7906
EXEC 14	0,7915	0,8605	0,2002	0,8052	0,7837	0,7943
EXEC 15	0,8017	0,864	0,2038	0,7924	0,8074	0,7998
EXEC 16	0,7833	0,8501	0,217	0,7826	0,7836	0,7831
EXEC 17	0,7937	0,865	0,2102	0,7868	0,7978	0,7922
EXEC 18	0,7949	0,8612	0,2048	0,7954	0,7946	0,795
EXEC 19	0,7892	0,8592	0,2091	0,792	0,7875	0,7897
EXEC 20	0,7909	0,8573	0,2098	0,7896	0,7916	0,7906
EXEC 21	0,7893	0,8572	0,2121	0,7868	0,7907	0,7887
EXEC 22	0,79	0,8594	0,2132	0,7844	0,7932	0,7888
EXEC 23	0,7951	0,8594	0,205	0,7942	0,7956	0,7949
EXEC 24	0,7938	0,859	0,2121	0,7834	0,8	0,7916
EXEC 25	0,7977	0,8639	0,202954	0,7966	0,7983	0,7974
EXEC 26	0,799	0,865	0,2064	0,7898	0,8046	0,7971
EXEC 27	0,7952	0,8639	0,205	0,7948	0,7954	0,7951
EXEC 28	0,7964	0,8609	0,2058	0,7926	0,7986	0,7956
EXEC 29	0,7997	0,8658	0,2028	0,7954	0,8023	0,7988
EXEC 30	0,7897	0,8576	0,2098	0,7904	0,7892	0,7898
Média (%)	79,47%	86,11%	20,69%	79,19%	79,64%	79,41%
DP (%)	0,50%	0,45%	0,52%	0,62%	0,63%	0,49%