Classificação de células tumorais em imagens histopatológicas utilizando Deep Learning

Breno C. Zukowski[1], **Lucas B. Figueira**[1]

[1]Faculdade de Tecnologia de Ribeirão Preto - (FATEC)
Ribeirão Preto, SP – Brasil

breno.marques@fatec.sp.gov.br, lucas.figueira@fatec.sp.gov.br

Abstract. *Breast Cancer diagnosis can be challenging even for well-trained doctors. Misleadings diagnosis inferred by doctors in training, and disagreement among histopathologists reports make harder to proceed an efficient patient care. With CAD (Computer Aided Diagnosis) it is possible to achive satisfactory results for physicians. This paper shows the use of a deep learning model that uses YOLOv5 to validate the effectiveness of state-of-the-art computer vision methods for detecting malignant tumors in medical images.*

Resumo. *O diagnóstico de câncer de mama pode ser desafiador e laborioso mesmo para profissionais bem treinados. Médicos inexperientes e discordância entre histopatologistas são causadores primários de diagnósticos errôneos que comprometem o tratamento de pacientes. Com sistemas computacionais de apoio a decisão que realizam classificação de objetos, é possível garantir melhores resultados para este processo. Este artigo demonstra a utilização de um modelo de aprendizagem profunda que utiliza YOLOv5 para validar a eficácia de métodos de visão computacional de estado da arte para detecção tumores malignos em imagens médicas.*

1. Introdução

O câncer de mama é a mais recorrente neoplasia maligna em mulheres ao redor do mundo. Apenas em 2020 foram realizados 2,3 milhões de diagnósticos e 685.000 mortes foram registradas (WHO, 2021). No Brasil, o cenário é similar: no mesmo ano a ordem de incidência estava prevista para cerca de 600.000 casos (INCA, 2018). A análise de imagens histológicas está entre os mais utilizados métodos de diagnóstico da atualidade. Porém, existem deficiências associadas ao método provenientes do trabalho humano desenvolvido para realizá-lo. Falhas estas, que podem levar a diagnósticos errados e agravamento do quadro de saúde do paciente em decorrência da falta de tratamento imediato. Mesmo quando bem sucedidos, a análise humana demanda uma grande carga de esforço e tempo que poderiam ser mitigados com auxílio de visão computacional e Deep Learning.

Segundo Tiezzi, Plotze e Figueira (2020), uma série de fatores podem ser descritos como métodos de predição de prognóstico, sendo atualmente utilizados no contexto clínico para determinação de tratamento, em especial para utilização de drogas antineoplásicas. Dentre eles destacam-se critérios clínicos, histológicos e utilização de marcadores tumorais. Abordando o critério histológico, temos o grau de diferenciação tumoral, baseado no sistema de escore de Nottingham (NGS), que, apesar de ser considerado um potente método de predição, recebe críticas em relação à sua baixa reprodutibilidade,

provavelmente devido ao seu caráter subjetivo e processamento pré-analítico da amostra. Dessa forma, gera ampla discordância entre histologistas que o aplicam, o que impacta diretamente no prognóstico do paciente e na decisão clínica de administrar ou não a quimioterapia sistêmica (TIEZZI; PLOTZE; FIGUEIRA, 2020). Alternativas de métodos com biologia molecular vêm sendo propostas para inferir com maior acurácia o estágio de agressividade da doença de forma a evitar desvios de diagnóstico. Entretanto, essas são técnicas de alto custo, inviáveis em muitas situações, principalmente em países subdesenvolvidos.

Atualmente, as técnicas de aprendizado de máquina vêm ganhando espaço em diversas áreas e aplicações. Na medicina, já são importantes métodos de auxílio ao diagnóstico de imagens radiológicas (HU et al., 2018). Diversos modelos computacionais têm sido desenvolvidos nos últimos anos utilizando aprendizagem profunda para concretizar sistemas de apoio ao diagnóstico. Grupos de pesquisa ao redor do mundo têm desenvolvido soluções de aprendizado de máquina utilizando técnicas diversas de *Deep Learning*, que, apesar de rápidas e geralmente acuradas, apenas oferecem mapas de calor e pontos de atenção, informações insuficientes para interpretação concreta e justificativa do diagnóstico oferecido pela máquina, o que não é adequado para sistemas de apoio à decisão médica (LI et al., 2021).

Vê-se, portanto, nas CNNs (do inglês, Convolutional Neural Network) de classificação de objetos em imagens, uma solução viável para análise de recortes específicos de tecido com a quantidade de informação e assertividade adequadas para auxílio ao diagnóstico médico. Pois a partir delas é possível segmentar e classificar regiões de interesse que evidenciam a presença de tumores malignos com eficiência.

O presente artigo propõe a utilização de um modelo de aprendizagem profunda para a segmentação de áreas de interesse e posterior classificação de imagens histológicas, dessa forma verificando a viabilidade de apoiar o diagnóstico de câncer de mama com este tipo de ténica.

2. Revisão bibliográfica

2.1. Introdução

Para justificar sua utilização é primeiro importante entender os conceitos que embasam as CNN's, técnicas estas utilizadas como base para a realização da detecção e classificação de objetos. Portanto entendemos como CNN: algoritmos de aprendizagem profunda especializados em processamento e classificação de imagens. Segundo (GOODFELLOW; BENGIO; COURVILLE, 2016) redes neurais convolucionais são um tipo especializado de rede neural para processamento de dados organizados topologicamente em grades, que através de operações matemáticas chamadas convoluções, são capazes de extrair características principais das entradas utilizando filtros (kernels), garantindo eficiência e redução de custos computacionais para a classificação.

Diferentemente de outros tipos de dados, imagens possuem a propriedade de *invariância de tradução* (AGGARWAL, 2018), ou seja, transmitem a mesma informação sobre o objeto independentemente das variações do contexto. No entanto, existem características espaciais, de luz, sombra, perspectiva, cenário, entre outras variáveis, que alteram significantemente a matriz de pixels que constituem a imagem do objeto de estudo.

Enquanto para um ser humano distinguir um objeto qualquer no espaço independente da posição, incidência de luz ou cenário em que este se encontra seja uma tarefa trivial, para um computador esta tem um custo elevado de execução. Sendo assim, é necessário que se utilize de técnicas que permitam a extração de características mais objetivas e com dados relevantes para a análise, evitando imprecisões que não contribuem para a classificação do objeto desejado.

Em geral, uma CNN apenas difere de uma rede neural densa totalmente conectada por apresentar camadas destinadas ao processamento das imagens que serão analisadas. Apresentando camadas de convolução, ativação, pooling, tal como apresentado na Figura 1 para finalmente enviar esses dados para uma rede neural densa que fará a classificação.

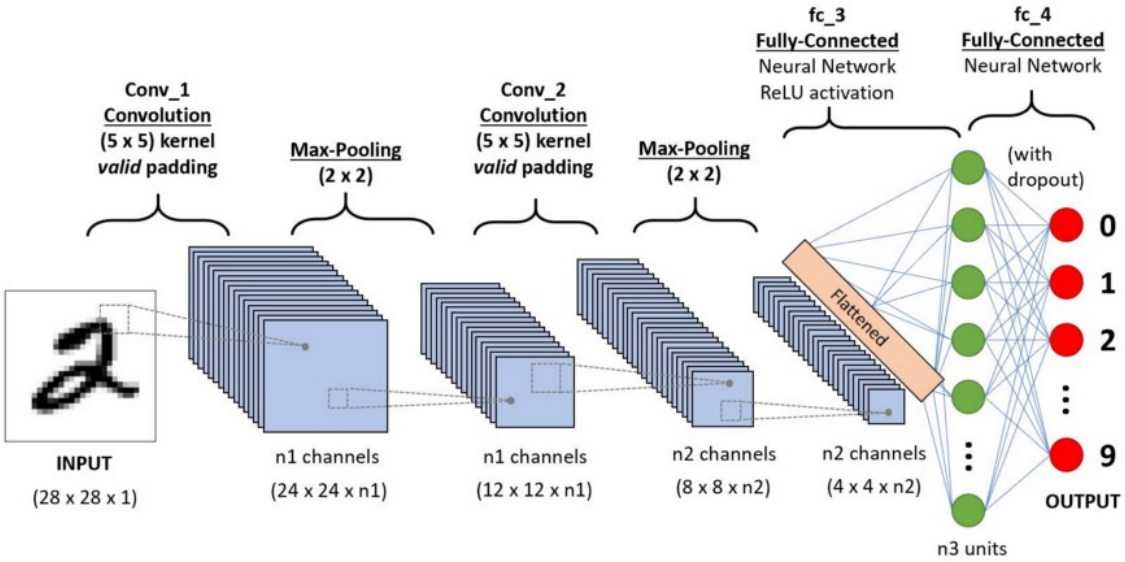

Figura 1. O fluxo básico de uma CNN e suas diversas camadas. Fonte: Introduction to Convolutional Neural Networks (CNN). Disponível em: https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/

2.2. Convolução

Em aplicações de aprendizagem de máquina a convolução é comumente interpretada como uma operação entre a imagem I e o kernel K, matrizes multidimensionais que ao serem convolucionadas resultarão em uma nova matriz que chamamos de *feature map* ou mapa de características. Quando tomamos por exemplo uma convolução de matrizes bidimensionais temos de levar em consideração certas propriedades matemáticas que não se traduzem bem para cenários práticos de aprendizagem profunda. Dessa forma, muitos *frameworks* de redes neurais implementam uma função similar chamada correlação cruzada (*cross-correlation*), que é em suma a mesma operação sem a rotação do kernel. Sendo definida pela seguinte função:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n)$$

Em termos gerais, basta que multiplique-se os valores das posições equivalentes do kernel nas coordenadas da imagem analisada, posteriormente soma-se todos os resultados para que se obtenha um valor único na posição definida (i, j) do *feature map*. É possível visualizar este processo com clareza na Figura 2.

O kernel é constituído por parâmetros que filtram certas características desejadas do *input*, uma rede neural convolucional pode apresentar diversas camadas de convolução para extração de características diferentes. As primeiras camadas em geral extraem características mais gerais, como bordas e contornos, enquanto as camadas posteriores extraem características mais específicas e abstratas.

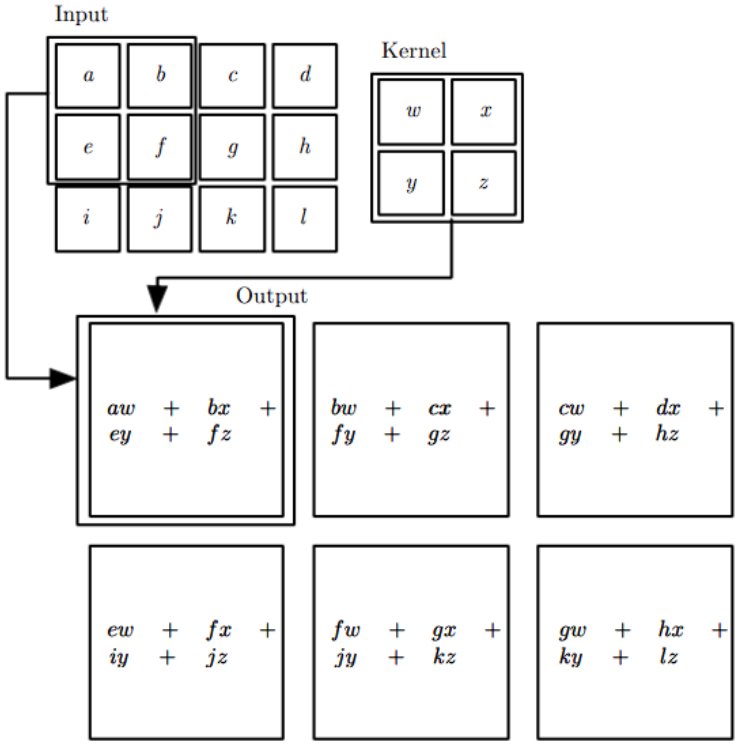

Figura 2. O processo de convolução sem rotação explicado graficamente. Imagem retirada do livro *Deep Learning* (GOODFELLOW; BENGIO; COURVILLE, 2016). Onde é possível perceber que o output é restringido apenas as posições compreendidas dentro da imagem por todos os parâmetros do kernel. São desenhadas caixas para indicar as saídas referentes a multiplicação e somatório das posições equivalentes do kernel à imagem.

2.3. Pooling

Apesar da efetividade de destaque de características provenientes da técnica de convolução, o *feature map* ainda destaca demasiadamente informações sobre localização espacial das *features* na saída, gerando uma variável extra a ser analisada (característica e posição) que prejudica a invariância de tradução da imagem. É mais relevante em geral para o modelo, compreender que a característica desejada está presente no objeto, do que sua posição especificamente (GOODFELLOW; BENGIO; COURVILLE, 2016). Portanto, em um fluxo comum de CNN é necessário que após o processo de convolução e aplicação da função de ativação (transformação não linear realizada ao longo do sinal

de entrada), seja realizada uma operação que permita suavizar a saída classificada. A esta técnica damos o nome de *pooling*.

A camada de *pooling* substitui a saída do *feature map* por uma estatística aproximada das informações desejadas para a rede neural densa, garantindo uma aproximação de invariância que permite melhores resultados para as operações realizadas nesse conjunto de dados (GOODFELLOW; BENGIO; COURVILLE, 2016).

Existem diversas funções de *pooling*, as duas mais comuns são: *Average Pooling* e *Max Pooling*. Seus respectivos algoritmos tem propósitos distintos, entretanto ambos cumprem o papel de filtrar o *feature map* de tal forma a reduzir a dimensionalidade da imagem e destacar características principais a serem processadas pela rede neural densa. Gerando um novo mapa de características que chamamos de *pooled feature map*.

2.4. Detecção unificada de objetos em tempo real com YOLO

Segundo Redmon et al. (2015) alternativas conhecidas de classificadores como R-CNN's (*Regions with Convolutional Neural Networks*) utilizam métodos para primeiro gerar potenciais caixas delimitadoras (*do inglês, bounding boxes*) em uma imagem e posteriormente rodar um classificador para as áreas segmentadas. Após as classificações, é aplicado ainda pós-processamento para refinar as delimitações, eliminar duplicações e possíveis imprecisões. Esse tipo de modelo, devido a sua complexidade, torna-se lento e de difícil otimização pois cada componente individual de sua composição deve ser treinado separadamente.

Com a proposta de ser um classificador de objetos de alta performance, YOLO propõe uma abordagem diferente das convencionais entregando velocidade e acurácia com custo computacional reduzido, além de uma arquitetura escalável devido a sua simplicidade. O modelo YOLO reformula o problema de detecção de objetos para um problema de regressão, utilizando apenas uma rede neural para criar *bounding boxes* e indicar as probabilidades dos objetos estarem presentes nessas imagens (REDMON et al., 2015).

Para tal, o sistema divide a imagem de entrada em um grid $S \times S$. Caso o centro da imagem caia em uma determinada célula, esta fica responsável por fazer a detecção daquele objeto, por sua vez esta prevê B caixas delimitadoras e valores de confiança para cada uma delas formalmente definidos por $Pr(Object) * IOU_{pred}^{truth}$. Se não existir objetos na respectiva célula o fator de confiança é determinado como zero, do contrário a confiança de predição é representada pelo IOU (*Intersection Over Union*) da *bounding box* predita em relação a qualquer caixa delimitadora tomada como verdade absoluta (REDMON et al., 2015).

Ademais, cada *bounding box* é composta por 5 valores: x, y, w, h e confiança (REDMON et al., 2015). Onde x, y são as coordenadas do centro da caixa delimitadora, w, h são a largura e altura relativas a imagem completa. Vale ressaltar que são preditas também C potenciais classes condicionais por célula e sua probabilidade é dada por $Pr(Class_i|Object)$ e apenas se prevê um único conjunto probabilidades de classe por célula do grid, independente de quantas *bounding boxes* foram demarcadas no processo (REDMON et al., 2015).

Durante o momento de testes são multiplicados as classes condicionais e as confianças individuais de cada caixa predita como visto em (1) o que dá os fatores de

confiança específicos de cada *bounding box* e por sua vez o fluxo pode ser compreendido na totalidade a partir da Figura 3.

$$Pr(Class_i|Object) * Pr(Object) * IOU_{pred}^{truth} = Pr(Class_i) * IOU_{pred}^{truth} \quad (1)$$

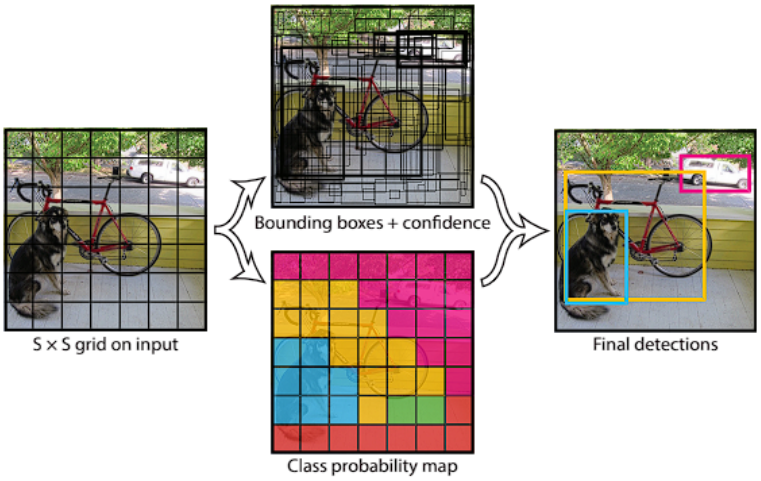

Figura 3. Imagem retirada do artigo de Redmon et al. (2015). Onde é possível ver com clareza o processo de separação da imagem em pequenas células em um grid SxS, a plotagem de B bounding boxes, o mapa de C probabilidades de classe e por fim um exemplo de detecção final.

3. Materiais e Métodos

Para concepção do modelo de detecção, utilizou-se uma seleção de imagens oferecidas por um especialista da área derivadas do banco de imagens TCGA [1], *dataset* este que possui informações clínicas e histopatológicas de 1098 pacientes com câncer de mama e suas respectivas imagens histológicas (TIEZZI; PLOTZE; FIGUEIRA, 2020).

Posteriormente, foram anotados por um especialista os tumores presentes em cada imagem que participaria do sub-conjunto de treino com as ferramentas *LabelImg* (TZUTALIN, 2015) e *Label Studio* (TKACHENKO et al., 2020-2022), estas são capazes de gerar anotações de classes em formato adequado para o treinamento como solicitado pela documentação oficial do YOLOv5 (JOCHER et al., 2022), que serão utilizadas como padrões de referência de verdade absoluta para o modelo. Vale ressaltar que a classe *tumor* é a única classe atribuída para os objetos relevantes nas imagens, não havendo distinção entre diferentes tipos de tumores malignos.

Para o processo de treinamento foram utilizadas 57 imagens que totalizam 1224 de objetos anotados. Este se deu em 250 épocas com lotes de 16 imagens, valores estes arbitrários com embasamento em exemplos da literatura. Devido ao tamanho reduzido do *dataset* para o treinamento, foi-se utilizado o modelo pré-treinado YOLOv5s (JOCHER et al., 2022) como é recomendado pela própria documentação oficial da arquitetura. Ao

[1] Disponível em: ⟨https://portal.gdc.cancer.gov/projects/TCGA-BRCA⟩ Acessado em: 19/08/2022

completar-se o processo, é gerado o novo modelo com pesos atualizados que pode ser utilizado para a detecção de tumores malignos.

Ademais, realizou-se a execução do script de validação em 14 imagens do subconjunto destinado para essa finalidade, para se obter as métricas que compõem a performance do sistema. Por fim, foram executados alguns testes para verificar a eficácia de anotação do modelo de forma visual em imagens selecionadas arbitariamente. Todo o fluxo pode ser observado na Figura 4.

Por fim, para a avaliação dos resultados foram utilizadas as métricas de *precision*, *recall*, *F1-score* e *mean average precision (mAP)* que serão visualizadas a partir de gráficos e relatórios gerados automaticamente pelo modelo no processo de treinamento e validação.

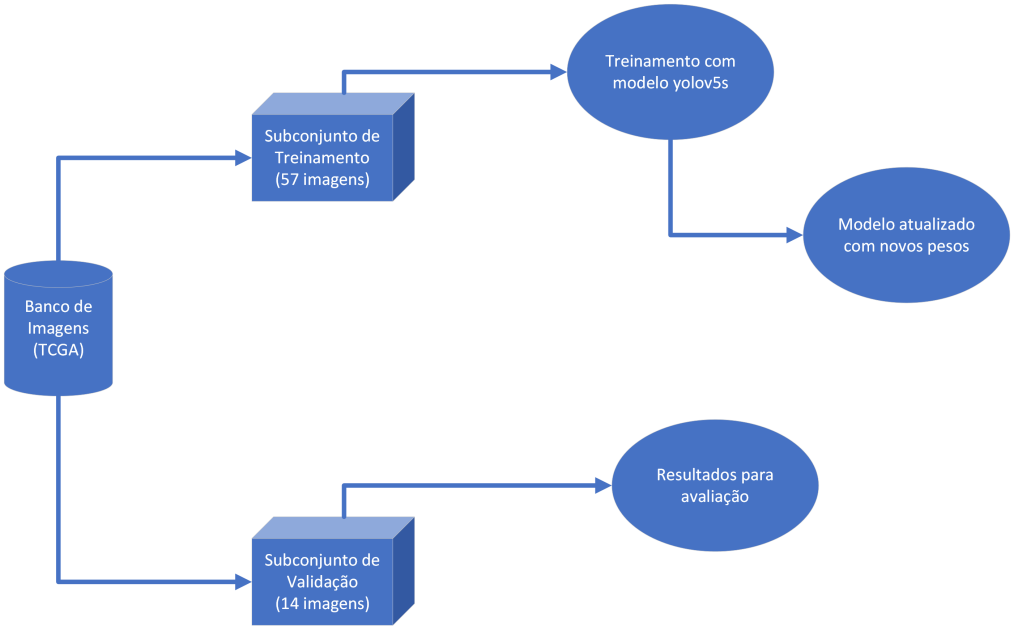

Figura 4. Fluxo do processo de treinamento e validação do modelo.

4. Resultados

Após a avaliação de resultados gerados conseguimos obter a acurácia de 81%. Como indicado na Figura 5, a precisão do modelo alcança o pico de 1.00 quando a confiança está em 0.882. Concomitantemente, o *recall* atinge seu valor mínimo de 0 quando o valor de confiança está em 0.96. Em relação ao mAP obtivemos os seguintes: mAP@0.5 com o valor 0.74 e mAP@0.5:0.95 em 0.383.

Ao se observar o balanço entre os gráficos de *precision* e *recall* entende-se que as detecções são mais abundantes até o nível de confiança 0.7, a partir daí o *recall* decai de forma expressiva indicando que o modelo já não mais acerta muito mais verdadeiros positivos. Esse entendimento pode ser validado observando o gráfico de *F1-score* que indica o balanço da média harmônica de *precision-recall* e atinge seu platô na confiança 0.72, posteriormente se aproxima do zero após a confiança de 0.8.

Como visto na Figura 6 o número de detecções se aproxima do *ground truth* no

limiar de confiança 0.7, abaixo disso a quantidade de detecções supera o número de tumores anotados. Entretanto, a partir de 0.8 detecções são realizadas chegando a zero em 0.9. O que nos permite inferir que apesar de não substituir o laudo de um patologista, existe viabilidade de papel de apoio do modelo para diagnósticos médicos, pois as quantidades de detecções corretas comparadas entre humano e máquina são de acurácia similares, se consideradas confianças inferiores a 80%.

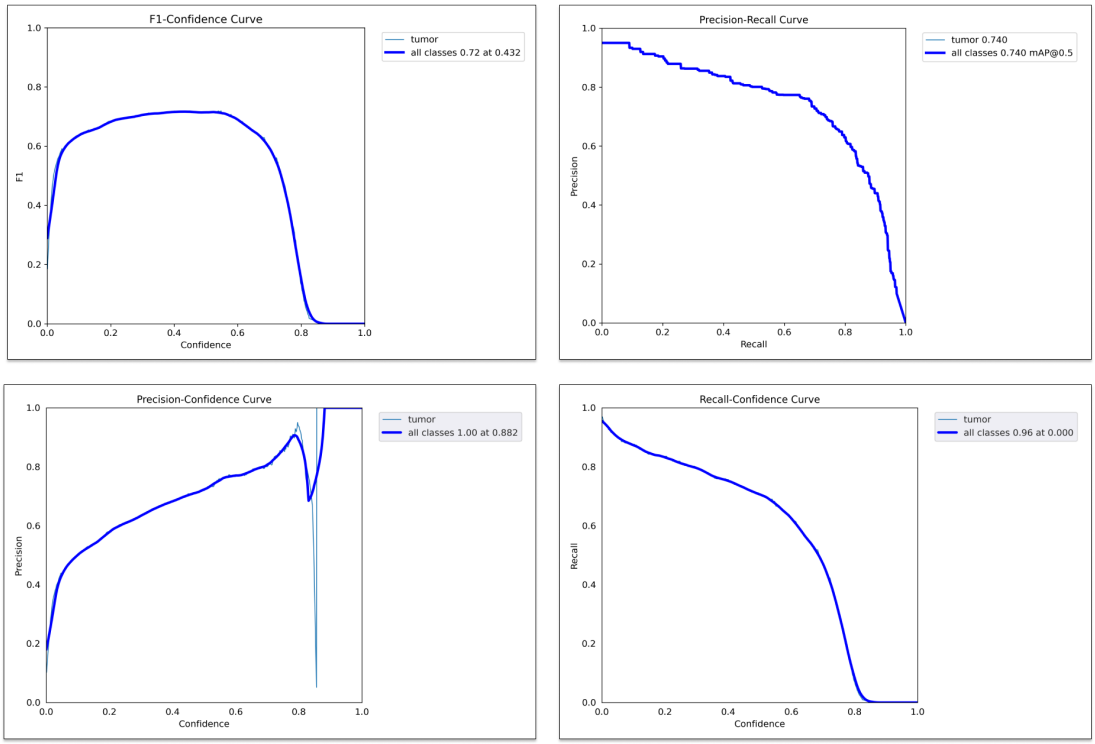

Figura 5. Gráficos de performances do modelo contendo respectivamente os gráficos relativos F1-Confiança, Precision-Recall, Precision-Confiança e Recall-Confiança.

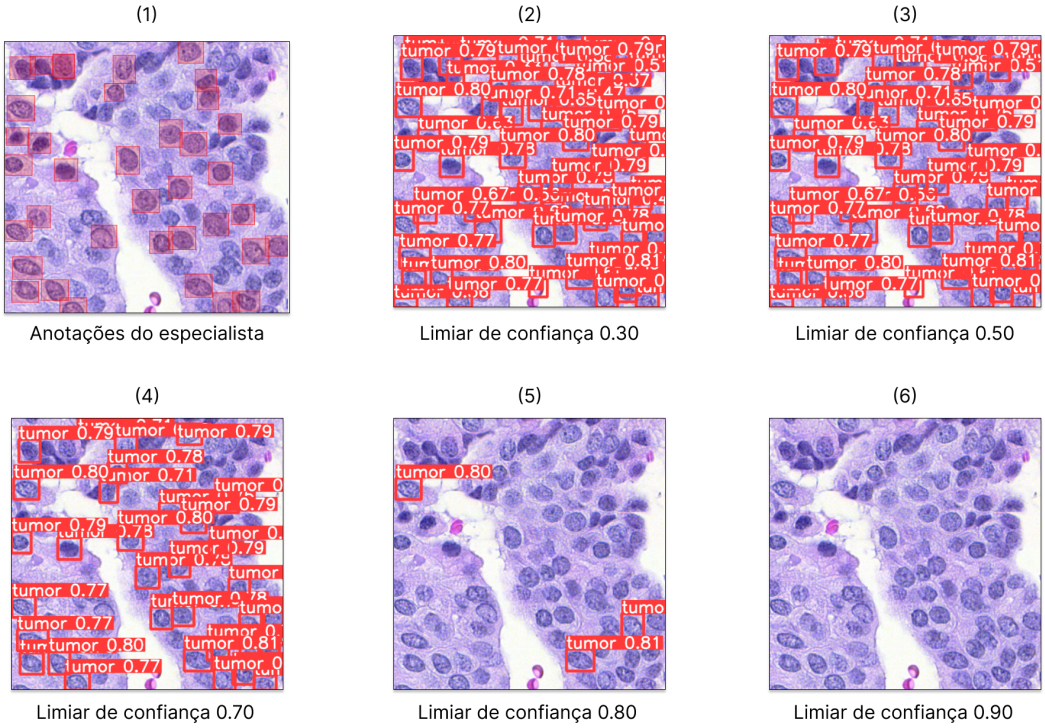

Figura 6. Anotações do especialista em (1) e nas imagems subsequentes as detecções realizadas pelo modelo com diferentes limiares de confiança para a detecção

5. Conclusões

Após a avaliação de resultados é possível perceber que existe viabilidade para a utilização de YOLOv5 com modelos pré-treinados como papel de apoio ao diagnóstico médico. Apesar de não substituir o laudo de um patologista, o modelo poderia estar presente no cotidiano de um consultório para apontar as imagens que merecem atenção especial do médico.

Ademais, é possível notar que existe espaço para evolução de performance se utilizadas maiores quantidades de dados de treinamento, principalmente para conseguir assertividades acima de 80% de confiança e maiores valores de *recall*.

Referências

AGGARWAL, C. C. *Neural Networks and Deep Learning*: A textbook. Cham: Springer, 2018. 497 p. ISBN 978-3-319-94462-3.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. ⟨http://www.deeplearningbook.org⟩.

HU, Z. et al. Deep learning for image-based cancer detection and diagnosis - a survey. *Pattern Recognition*, v. 83, p. 134–149, 2018. ISSN 0031-3203.

INCA. *INCA estima que haverá cerca de 600 mil casos novos de câncer em 2018*. 2018. Acessado em 05 de Fev. de 2022. Disponível em: ⟨https://www.inca.gov.br/imprensa/inca-estima-que-havera-cerca-de-600-mil-casos-novos-de-cancer-em-2018⟩.

JOCHER, G. et al. *ultralytics/yolov5: v6.2 - YOLOv5 Classification Models, Apple M1, Reproducibility, ClearML and Deci.ai integrations*. Zenodo, 2022. Disponível em: ⟨https://doi.org/10.5281/zenodo.7002879⟩.

LI, B. et al. Classifying breast histopathology images with a ductal instance-oriented pipeline. In: IEEE. *2020 25th International Conference on Pattern Recognition (ICPR)*. [S.l.], 2021. p. 8727–8734.

REDMON, J. et al. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015. Disponível em: ⟨http://arxiv.org/abs/1506.02640⟩.

TIEZZI, D. G.; PLOTZE, R.; FIGUEIRA, L. B. Deep learning como sistema de auxílio diagnóstico e classificação do câncer de mama. *I Workshop de Tecnologia da Fatec Ribeirão Preto*, v. 1, n. 1, 2020.

TKACHENKO, M. et al. *Label Studio: Data labeling software*. 2020–2022. Open source software available from https://github.com/heartexlabs/label-studio. Disponível em: ⟨https://github.com/heartexlabs/label-studio⟩.

TZUTALIN. *LabelImg*. 2015. Free Software: MIT License. Disponível em: ⟨https://github.com/tzutalin/labelImg⟩.

WHO. *Breast cancer*. 2021. Acessado em 05 de Fev. de 2022. Disponível em: ⟨https://www.who.int/news-room/fact-sheets/detail/breast-cancer⟩.