

Predição de lesões celulares em imagens de citologia convencional usando redes neurais convolucionais

Lucas A. Freitas¹, Débora N. Diniz¹, Marcone J. F. Souza¹, Cláudia M. Carneiro², Daniela M. Ushizima³, Fátima N. S. de Medeiros⁴, Andrea G. C. Bianchi¹

¹Departamento de Computação,
Universidade Federal de Ouro Preto (UFOP), Ouro Preto (MG), Brasil

²Departamento de Análises Clínicas,
Universidade Federal de Ouro Preto (UFOP), Ouro Preto (MG), Brasil

³Divisão de Matemática Aplicada e Pesquisa Computacional,
Lawrence Berkeley National Laboratory (LBNL), Berkeley (CA), Estados Unidos

⁴Departamento de Engenharia de Teleinformática,
Universidade Federal do Ceará (UFC), Fortaleza, CE, Brasil

{lucas.freitas2, debora.diniz}@aluno.ufop.edu.br, marcone@ufop.edu.br

dushizima@lbl.gov, fsombra@ufc.com.br, andrea@ufop.edu.br

Abstract. *This article presents a novel methodology based on deep learning for detecting cervical lesions in Pap smear samples. The proposed model uses nucleus location information and performs cropping around it using different dimensions, without the need for image segmentation. Several CNN models were developed and trained using real cervical cell images. The results showed that the model achieved a satisfactory accuracy of 0.94 using a box size of 70x70 without needing image segmentation. The proposed methodology can assist cytopathologists in improving diagnosis and laboratory results quality, contributing to cervical cancer prevention.*

Resumo. *Este artigo apresenta uma nova metodologia baseada em aprendizado profundo para detectar lesões cervicais em amostras de exame de Papanicolau. O modelo proposto utiliza informações de localização do núcleo e realiza recortes em torno dele usando diferentes dimensões, sem a necessidade de segmentação da imagem. Vários modelos de CNN foram desenvolvidos e treinados com imagens reais de células cervicais. Os resultados mostraram que o modelo atingiu uma acurácia satisfatória de 0,94 usando o tamanho de caixa de 70x70 sem a necessidade de segmentar imagens. Acredita-se que essa metodologia possa auxiliar os citopatologistas na melhoria do diagnóstico e na qualidade dos resultados dos laboratórios, contribuindo para a prevenção do câncer de colo do útero.*

1. Introdução

De acordo com a Organização Mundial de Saúde [Who 2022], em 2018 o câncer do colo do útero foi o terceiro tumor mais frequente na população feminina mundial, com uma média de 530.000 novos casos da doença. Ainda segundo esse relatório, esse tipo de

câncer foi responsável por 265 mil óbitos anuais, tendo sido a quarta causa mais frequente de morte por câncer em mulheres. No entanto, um fato positivo é que essa neoplasia tem desenvolvimento lento, o que eleva consideravelmente a taxa de cura quando essas lesões são diagnosticadas precocemente pelo exame citopatológico ou Papanicolau.

O exame de Papanicolau permite identificar alterações nas células descamadas do colo do útero com base em vários estudos citomorfológicos. Os critérios são baseados em características nucleares, sendo um exemplo o aumento nuclear, a irregularidade da membrana, a hipercromasia, e a relação do tamanho do núcleo e citoplasma. Com o passar dos anos, o exame de Papanicolau se tornou a principal forma de diagnóstico em países em desenvolvimento, principalmente pelo fato de os outros métodos de exame serem extremamente caros.

Normalmente um citopatologista costuma avaliar até 300.000 células cervicais em um único esfregaço. Além disso, dependendo da sua nacionalidade, ele pode chegar a avaliar 100 esfregaços por dia. No Brasil, por exemplo, a recomendação de análises diárias é de no máximo 70 esfregaços. Porém, nem sempre esse limite é respeitado, o que pode ocasionar erros no resultado do exame. Uma investigação conduzida desde antes de 1990 mostra que as taxas de falsos-negativos, que representa os casos de células anormais classificados como normais, podem variar entre 2% a 62% [Sung et al. 2021] nos exames de Papanicolau, o que é considerado um problema grave.

Neste artigo é proposto o desenvolvimento de um modelo de classificação de células cervicais em imagens com múltiplas células, com o objetivo de reconhecer três classes distintas: células com lesão, células sem lesão e não células. Para isso, são adotadas estratégias previamente descritas em estudos anteriores, notadamente em [Diniz et al. 2021a] e [Isidoro 2021]. O trabalho de [Diniz et al. 2021a] foi utilizado como base para a elaboração do modelo de treinamento da rede neural, enquanto o de [Isidoro 2021] foi fonte de inspiração para a criação de regiões de interesse na imagem a ser analisada.

Os objetivos deste artigo consistem em: (a) detectar regiões com lesão sem o uso de segmentação ou delimitações de regiões na imagem; (b) investigar diferentes abordagens para o treinamento do modelo, utilizando regiões centralizadas ou não no núcleo e diferentes tamanhos de regiões; (c) apresentar regiões com maior probabilidade de lesão; e, (d) realizar a predição de lesões em imagens de citologia convencional. Com esses objetivos, espera-se contribuir para o aprimoramento dos métodos de análise de células cervicais, com potenciais impactos positivos na detecção precoce de lesões cervicais e prevenção do câncer de colo do útero.

2. Trabalhos Relacionados

Em [Diniz et al. 2021b] são utilizados diversos algoritmos tradicionais de aprendizado de máquina para realizar uma classificação hierárquica em até 6 classes. A detecção é baseada na análise das características dos núcleos das células das imagens do esfregaço de Papanicolau. Dentre os algoritmos utilizados estão presentes a *Árvore de Decisão*, *K-Nearest Neighbors* (k-NN), *Random Forest* e *Métodos de Ridge*. Os autores concluíram que o melhor resultado se obtém ao usar o *Random Forest* como classificador.

[Diniz et al. 2022] descrevem o desenvolvimento de uma ferramenta chamada CEA (*Cytopathologist Eye Assistant*) para auxiliar citopatologistas na análise de imagens

de esfregaços cervicais. A ferramenta utiliza uma combinação de algoritmos YOLOv5s para detecção de células e EfficientNetB0s para classificação. Os resultados mostram que a ferramenta alcançou uma precisão de 0,726, uma revocação de 0,906 e uma F1-score de 0,805 ao considerar células individuais, e uma precisão de 0,975, uma revocação de 0,992, uma acurácia de 0,970 e uma F1-score de 0,983 para a classificação de imagem como um todo. A ferramenta foi bem recebida pelos patologistas que a testaram e pode ser integrada em rotinas de análise de imagens de esfregaços cervicais para melhorar a qualidade dos resultados.

[Gómez et al. 2017] compararam vários algoritmos para classificação do câncer de colo de útero, separando as células em duas classes: normais e anormais. Eles seguiram o seguinte fluxo: aquisição de dados, pré-processamento, treinamento e seleção do modelo. O banco de dados utilizado foi fornecido pelo hospital da Herley University, que tinha disponível 917 imagens, das quais 242 eram células normais e 675 anormais. Os autores extraíram 20 características morfológicas das imagens, cada uma com seu peso e concluíram que as combinações de algoritmos Bagging + MultilayerPerceptron e AdaBoostM1 + LMT (*Logistic model tree*) foram as melhores analisadas por eles.

[Mariarputham and Stephen 2015] relataram que a tarefa de classificação precisa das imagens do esfregaço de Papanicolau podia ser melhorada através da seleção de características específicas bem definidas e da seleção do melhor classificador. O sistema apresentado envolvia a preparação dos esfregaços de Papanicolau através de imagens capturadas por microscópio óptico passadas para um pré-processamento onde ocorre a segmentação do núcleo do citoplasma. Logo após, são extraídas características sobre a textura das imagens segmentadas. Após a realização dessas etapas, é realizada a classificação em 7 possíveis classes utilizando o *Support Vector Machine* (SVM). Os autores ressaltam que não existe um conjunto exclusivo de características capazes de fornecer diagnósticos eficientes para todas e quaisquer classes de um sistema de classificação proposto. No entanto, o SVM foi reconhecido como uma solução eficiente para o problema proposto.

[Isidoro 2021] propõe uma classificação automática de recortes de células cervicais com base em características não geométricas. A classificação é feita utilizando métodos tradicionais na sua arquitetura interna, seguindo o seguinte fluxo: as imagens são retiradas do banco de dados do *Center for Recognition and Inspection of Cells* (CRIC) e recortadas em sub-imagens com dimensão fixa de 50x50. As sub-imagens passam por um extrator de características e são aplicadas diversas metodologias para conseguir extrair o máximo de informação daquela imagem. Em seguida é feito o treinamento utilizando o SVM. Testes foram realizados aplicando-se a validação cruzada, sendo 20% do total das imagens utilizadas somente para teste. Por fim, foram reportados os valores de várias métricas usadas para avaliar a eficiência da proposta. O autor concluiu que a abordagem proposta teve ótimos resultados.

3. Desenvolvimento

3.1. Base de Dados Cric Cervix

A base de dados selecionada para a extração das imagens deste trabalho, juntamente com os núcleos celulares, é a CRIC [Rezende et al. 2021], uma coleção de imagens de células cervicais. A base de dados contém 11.534 células classificadas, que foram obtidas em

400 imagens de exame de Papanicolau. Na Figura 1a temos um exemplo de uma imagem extraída dessa base de dados. Os dados são compostos por imagens multicelulares e pelo mapeamento da posição x e y de todos os núcleos/células existentes nas imagens e suas respectivas classificações. A posição marcada pelo profissional encontra-se na região central do núcleo celular. As células dessa coleção são classificadas em seis classes baseadas na nomenclatura Bethesda de lesões [Pangarkar 2022], representadas na Figura 1b:

1. Negativo para lesão intraepitelial ou malignidade (NILM);
2. Células escamosas atípicas de significado indeterminado, possivelmente não neoplásicas (ASC-US);
3. Lesão intraepitelial escamosa de baixo grau (LSIL);
4. Células escamosas atípicas, não pode excluir lesão de alto grau (ASC-H)
5. Lesão intraepitelial escamosa de alto grau (HSIL);
6. Carcinoma de células escamosas (SCC)

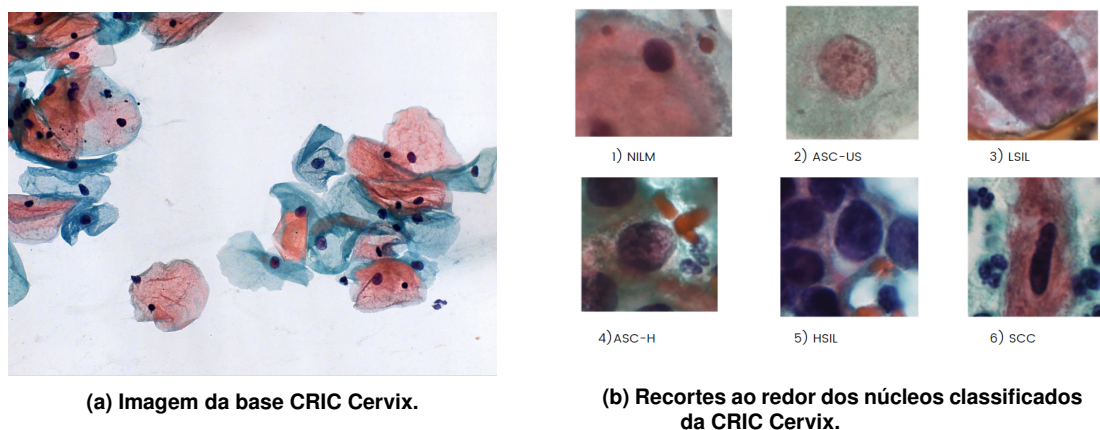


Figura 1. CRIC Cervix [Rezende et al. 2021].

Apesar de a base de dados possuir seis classes diferentes, neste trabalho ela é separada em apenas duas classes: células sem lesão ou normal (NILM) com 6.770 células; e células com lesão ou anormal (LSIL, HSIL, ASC-H, ASC-US, NILM) totalizando 4.755 células. A base é composta por imagens contendo um número variável de núcleos, o que nos leva a ter mais núcleos normais do que anormais, deixando assim a base desbalanceada. A distribuição do número de lesões celulares em cada classe está disponível na Tabela 1.

Tabela 1. Distribuição de núcleos por classe na CRIC Cervix.

Classes	Categoria	Tipo de célula	Total de lesões
1	Anormal	LSIL	1.360
2	Anormal	HSIL	1.703
3	Anormal	ASC-H	925
4	Anormal	ASC-US	606
5	Anormal	SCC	161
6	Normal	NILM	6.779

3.2. Pré-processamento

Conforme comentado, a marcação de classificação proporcionada pela base de dados CRIC Cervix era um único ponto centrado em cada núcleo celular. Esta abordagem, apesar de simples e de fácil marcação pelo profissional, apresenta algumas limitações pois não temos exatidão da região do núcleo ou do citoplasma celular. Desse modo, temos que adotar estratégias de escolha do que será enviado para o modelo de redes neurais para o treinamento de lesões celulares.

No pré-processamento foram adotadas duas estratégias diferentes nas imagens para geração do conjunto de treino e teste, gerando dois tipos de imagens: as com o núcleo centralizado, e as com o núcleo descentralizado. A base de dados utilizada para o treinamento e o teste incluiu a soma dos resultados das duas estratégias.

3.2.1. Núcleo Centralizado

Para cada imagem da base CRIC Cervix é mapeada a posição x e y da célula nela presente (e sua classificação), tendo o núcleo como ponto de referência. Para a geração dos recortes com o núcleo centralizado, foram criadas caixas de quatro tamanhos distintos (30x30, 50x50, 70x70 e 90x90), considerando o ponto x e y na região central de cada caixa. Dessa maneira garante-se que no centro do recorte esteja sempre o núcleo celular. Realizamos uma investigação do tamanho da caixa em função da qualidade da classificação, propondo algumas variações de tamanho. No trabalho proposto por [Diniz et al. 2021b] foi utilizado um valor fixo para o tamanho da caixa que continha a região do núcleo. Caso haja núcleos na borda da imagem, eles são descartados.

A Figura 2a exemplifica o processo de geração dos recortes utilizando um tamanho de caixa 90x90. As caixas verdes apresentam as células classificadas como sem lesão e os quadrados vermelhos são as células classificadas como com lesão pelos especialistas. Já a Figura 2b representa um exemplo de uma mesma célula para os diferentes tamanhos de caixa utilizados. Após a criação dos recortes, para cada tamanho de imagem 30x30,

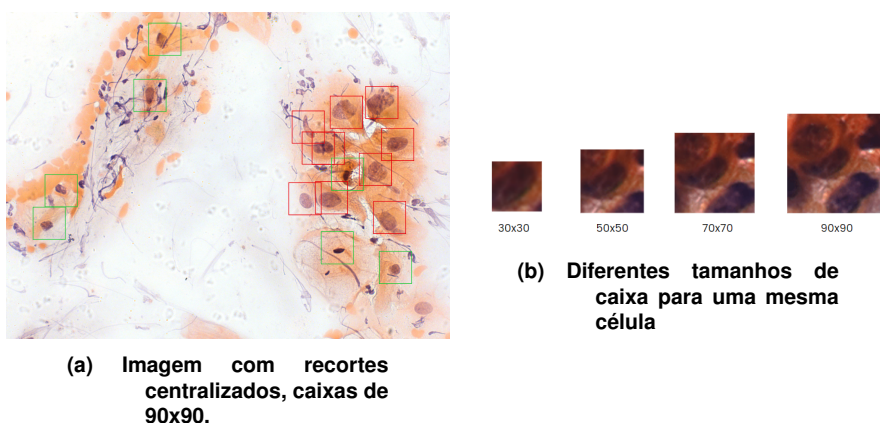


Figura 2. Geração de recortes centralizados.

50x50, 70x70 e 90x90 foram gerados os mesmos números de imagens, pois ao usar a abordagem centralizada todos os núcleos foram mapeados e utilizados, justamente por nunca ter informações no limite do recorte. O número total de recortes classificados com

lesão e sem lesão foram, respectivamente 4.755 e 6.779. Para a elaboração do modelo final adotamos a estratégia 70-30, que consiste em separar 70% das imagens para treino e 30% das imagens para teste.

3.2.2. Núcleo Descentralizado

Para os recortes com o núcleo descentralizado, foi adotada uma abordagem diferente, pois a ideia principal era ter imagens onde o núcleo não estivesse apenas no centro da imagem. Para isso, percorremos toda a imagem do exame de Papanicolau, recortando a imagem de forma sequencial de acordo com o tamanho de caixa desejado. Um exemplo deste processo é apresentado na Figura 3a. Para cada caixa era consultado se existia um núcleo em seu interior, e, se ele tinha ou não lesão. As caixas verdes representam os recortes que possuem núcleo e que são normais, e as vermelhas recortes que possuem núcleo com lesão. As caixas que não possuíam núcleo dentro delas eram salvas com o rótulo de não célula (em amarelo), pois a classe não célula também foi incluída no modelo.

Em alguns casos, o núcleo da célula localizava-se no limite do tamanho da caixa, para que isso não acontecesse e as informações não fossem muito prejudicadas, adotamos uma estratégia de fronteira, na qual a caixa só é armazenada se o núcleo estiver a pelo menos 10 *pixels* de distância das bordas da caixa.

A Figura 3a representa o processo de geração das imagens descentralizadas. Nesta figura, as caixas amarelas representam as imagens classificadas como não célula, as vermelhas como com lesão e as verdes como sem lesão. Na Figura 3b é mostrado o resultado final da estratégia da fronteira. É possível observar que com essa estratégia várias caixas deixam de ser consideradas por não respeitarem o limite de 10 *pixels* estipulado. Neste exemplo não foi mapeada a classe não célula apenas para facilitar a visualização. Dessa maneira, apenas as classes com lesão (vermelha) e sem lesão (verde) foram apresentadas.

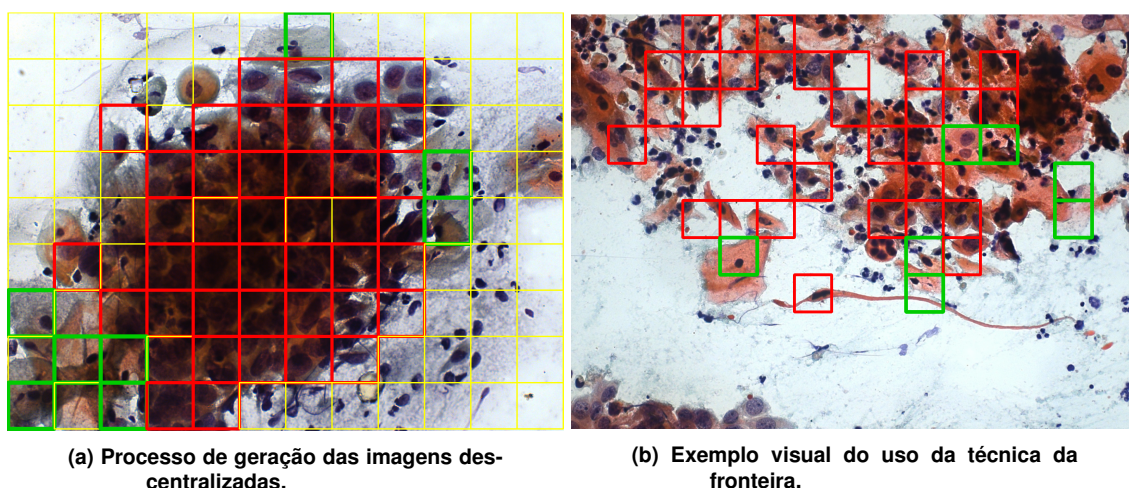


Figura 3. Geração recortes descentralizados.

Diferentemente da geração das imagens centralizadas, nas imagens descentralizadas houve variações na quantidade de caixas contendo células. Isso aconteceu devido à

estratégia da fronteira, já que cada caixa varia o tamanho e isto influencia na quantidade de recortes gerados.

Após gerados os recortes descentralizados para todos os tamanhos de caixa, foi preciso balancear o banco de dados, pois cada classe possuía um número diferente de instâncias, sendo a classe com lesão a limitadora. Dessa forma, foi somado a classe com lesão centralizada e com lesão descentralizada, gerando o número total de recortes para a classe núcleo com lesão. Este número foi usado para limitar as outras classes; sendo assim, algumas instâncias foram removidas a fim de tornar o banco de dados balanceado. A Tabela 2 representa o número total de instâncias para a classe 70x70 após o balanceamento.

Tabela 2. Distribuição final das classes para tamanho de caixa 70x70.

Conjunto	Com Lesão	Sem Lesão	Não Célula
Treinamento	4.867	4.886	4.893
Teste	2.086	2.094	2.097
Total	6.953	6.980	6.890

3.3. Arquitetura das Redes Neurais Convolucionais

Com a evolução da inteligência artificial, diferentes arquiteturas de rede neural convolucional têm sido utilizadas para classificar imagens. Como este trabalho visa a investigação do uso de diferentes modelos para a classificação das imagens, foram utilizados quatro modelos diferentes, sendo eles (*EfficientNetB0*, *MobileNet*, *XceptionNet* e *Inception V3*). A nossa escolha foi baseada nos modelos utilizados em [Diniz et al. 2021a], pois temos o interesse em comparar os resultados obtidos, já que serão utilizados os mesmos modelos variando apenas o pré-processamento.

4. Resultados

4.1. Descrição dos Experimentos

Para o desenvolvimento deste trabalho, os experimentos foram executados por meio da máquina virtual do *Google Colaboratory*. As particularidades do *hardware* estão presentes na Tabela 3. Por sua vez, a implementação dos modelos das redes neurais convolucionais foram desenvolvidas utilizando a biblioteca *TensorflowKeras*.

Tabela 3. Especificação do Hardware utilizado.

Parâmetros	Especificação
GPU	Nvidia K80
Memória da GPU	12 GB
Performance	4.1 TFLOPS
Número de núcleos da GPU	2
Memória RAM	12 GB
Memória em Disco	358 GB

Para todos os modelos com seus diferentes tamanhos de recortes, foram definidas algumas variáveis padrão. Todos os experimentos foram executados com 40 épocas,

tendo como função de ativação o *Softmax* e como otimizador a função de Adam. Como mencionado, a base de dados foi dividida em 70% para treino e 30% para teste e o tamanho do lote foi fixado em 32 para todos os experimentos. Todos esses valores foram definidos para seguirem o mesmo padrão adotado em [Diniz et al. 2021a].

Tamanho de recorte 30x30

No tamanho de recorte 30x30, apenas o modelo *EfficientNetB0* foi capaz de ser executado com imagens deste tamanho devido à limitação de tamanho mínimo de entrada dos outros modelos. Na Tabela 4 estão apresentados os valores das métricas encontradas pelo modelo. A Figura 4a representa a matriz de confusão do modelo, que representa a distribuição das classes que o modelo inferiu.

Tabela 4. Valores das métricas do modelo *EfficientNetB0* para o recorte de 30X30.

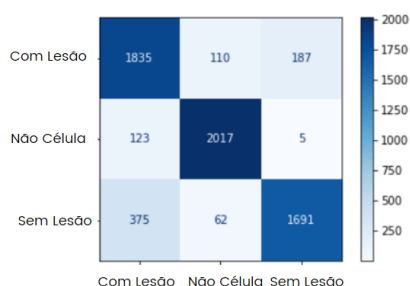
Arquitetura	Precisão	Acurácia	F1-Score	Revocação	Especificidade
EfficientNetB0	0,87	0,91	0,87	0,87	0,93

Tamanho de recorte 50x50

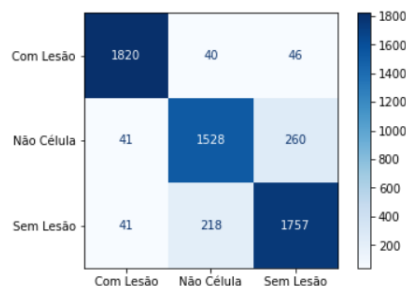
Para o tamanho de recorte 50x50, dois modelos estão aptos a processar as imagens, sendo eles: *EfficientNetB0* e *MobileNet*. Na Tabela 5 estão reportados os valores das métricas encontradas. A Figura 4b mostra a matriz de confusão do modelo *EfficientNetB0*, que obteve resultados ligeiramente melhores que o do modelo *MobileNet*, com quase todas as métricas com valores cerca de 1% maiores e acurácia com o mesmo valor.

Tabela 5. Valores das métricas dos modelos para o recorte de 50X50.

Arquitetura	Precisão	Acurácia	F1-Score	Revocação	Especificidade
EfficientNetB0	0,90	0,93	0,90	0,90	0,95
MobileNet	0,89	0,93	0,89	0,89	0,94



(a) Matriz de confusão do modelo *EfficientNetB0* para o tamanho de caixa 30x30.



(b) Matriz de confusão do modelo *EfficientNetB0* para o tamanho de caixa 50x50.

Figura 4. Matrizes de confusão dos recortes 30x30 e 50x50.

Tamanho de recorte 70x70

No recorte de 70x70, todos os modelos estão aptos a processar as imagens e seus resultados estão reportados na Tabela 6. A Figura 5a mostra a matriz de confusão do modelo *EfficientNetB0*, o qual obteve novamente o melhor resultado. Nota-se que todos os modelos retornaram valores muito próximos, variando no máximo 2,3% entre o melhor e o pior modelo.

Tabela 6. Resultados das métricas dos modelos para o recorte de 70x70.

Arquitetura	Precisão	Acurácia	F1-Score	Revocação	Especificidade
EfficientNetB0	0,90	0,94	0,90	0,90	0,95
MobileNet	0,89	0,92	0,89	0,89	0,94
InceptionV3	0,88	0,92	0,88	0,88	0,94
XceptionNet	0,90	0,93	0,90	0,90	0,95

Tamanho do recorte 90x90

Por fim, para o último recorte usamos todos os modelos propostos. Na Tabela 7 estão reportados os valores de todas as métricas. A Figura 5b mostra a matriz de confusão do modelo *XceptionNet*, que nesse tamanho de caixa obteve os melhores resultados. Percebe-se que, nesse tamanho de caixa, os modelos variaram mais entre si, com uma diferença de 18,5% do melhor modelo para o pior. Na tabela 7 está evidenciado também o melhor resultado de [Diniz et al. 2021a] que utilizou esse mesmo tamanho de recorte. Porém, vale a pena ressaltar que se trata de uma abordagem diferente, pois seus recortes foram apenas centrados no núcleo e não contemplavam uma classe não célula, classificava apenas regiões com núcleo.

Tabela 7. Resultados das métricas dos modelos para o recorte de 90x90.

Arquitetura	Precisão	Acurácia	F1-Score	Revocação	Especificidade
EfficientNetB0	0,86	0,91	0,86	0,86	0,93
MobileNet	0,87	0,92	0,87	0,87	0,94
InceptionV3	0,81	0,87	0,81	0,81	0,90
XceptionNet	0,89	0,93	0,89	0,89	0,95
Ensemble¹	0,96	0,96	0,96	0,96	0,96

¹ [Diniz et al. 2021a]

4.2. Discussão dos Resultados

Nesta seção discutimos os resultados encontrados pelos diferentes modelos e diferentes tamanhos de caixa. Nota-se que os melhores resultados foram encontrados com o **recorte de 70x70** com a utilização do modelo *EfficientNetB0*.

Os resultados indicam que os modelos retornaram melhores resultados à medida que os recortes foram se tornando maiores. Porém, ao chegar ao tamanho de caixa 90x90, os modelos se diferenciaram mais entre si, encontrando resultados variados e, em média, todos os modelos obtiveram piores resultados em comparação com o tamanho de caixa 70x70.

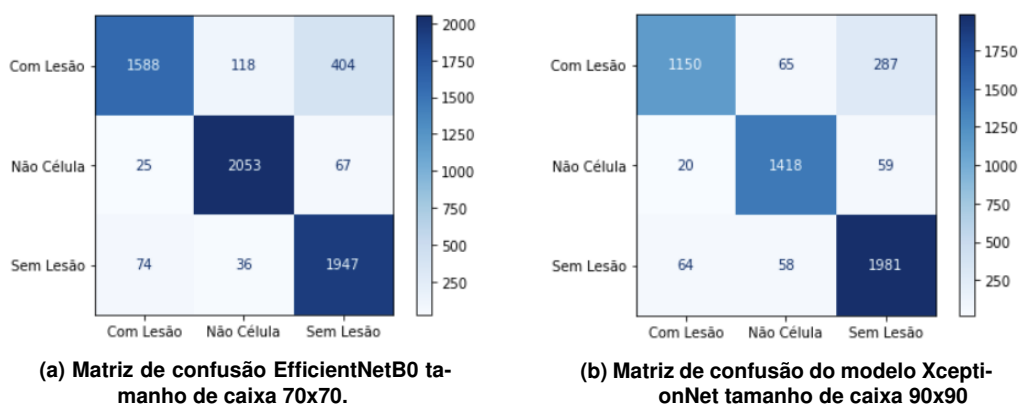


Figura 5. Matrizes de confusão dos recortes 70x70 e 90x90.

Uma hipótese é que os recortes muito pequenos representam muito pouca informação do núcleo, o que pode dificultar o processo de treinamento e as predições. Por outro lado, recortes muito grandes podem agregar informações além do núcleo e outras estruturas celulares, assim como sobreposição.

Ao analisar as matrizes de confusão geradas pelos modelos, constatou-se que o recorte de tamanho 30x30 apresentou o maior número de erros ao classificar a classe sem lesão como com lesão. Já no recorte de tamanho 50x50, o maior número de resultados incorretos foi observado na classificação da classe não célula como sem lesão. Por fim, tanto no tamanho 70x70 quanto no tamanho 90x90, foi observado um maior número de erros na classificação da classe com lesão como sem lesão. Esses resultados indicam a importância da escolha adequada do tamanho do recorte em tarefas de classificação de imagens, destacando a necessidade de avaliar diferentes abordagens para obter um modelo mais preciso e confiável.

Dessa maneira, para o desenvolvimento do sistema de detecção de lesões, foi utilizado o modelo e o recorte que proporcionou melhores resultados. Sendo assim, foi utilizado o modelo *EfficientNetB0* com recorte 70x70.

4.3. Predição do Sistema

Esta seção mostra como o modelo se comportou para predições em imagens desconhecidas, evidenciando como ele classificou as diferentes caixas dentro das imagens.

Para realizar a predição utilizamos a mesma lógica da geração de recortes descentralizados. Sendo assim, a primeira coisa a se fazer foi redimensionar a imagem original para um tamanho que fosse possível tratar caixas de 70x70 por toda a imagem, sem deixar de cobrir nenhuma parte da imagem.

A imagem de entrada foi recortada e esses recortes serviram de entrada para a inferência do modelo. Caso ele classificasse como lesão, o recorte era desenhado como um quadrado em vermelho na imagem original, um quadrado verde para não lesão, seguindo o mesmo padrão adotado pela base de dados CRIC. Para os casos concluídos como não célula não foi desenhado nenhum quadrado a fim de facilitar a visualização das células e não poluir a imagem com muita informação. A Figura 6a mostra a verdadeira classificação de uma imagem presente na base de dados CRIC. Já a Figura 6b ilustra a mesma imagem da figura anterior classificada pelo modelo proposto.

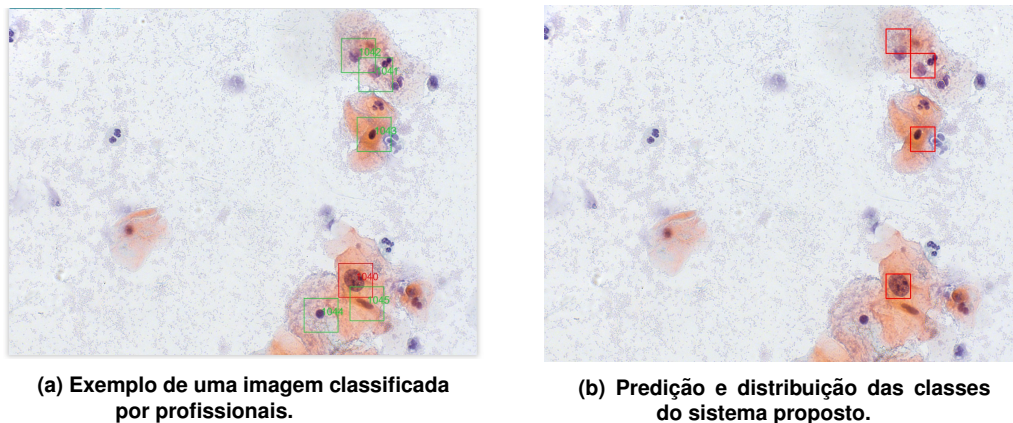


Figura 6. Geração recortes descentralizados.

Ao comparar as duas figuras percebemos que o modelo proposto conseguiu identificar muito bem as células, deixando de mapear apenas duas delas. Ele também obteve um bom desempenho em diferenciar célula de não célula. Porém, de todas as quatro células que ele reconheceu, ele concluiu todas como com lesão, sendo que somente uma era realmente classificada como com lesão, e as outras três como não lesão. Dessa forma, percebemos que o modelo se confundiu na hora de concluir o tipo de lesão da célula. Em [Diniz et al. 2021a] tivemos uma resposta similar, também em ambiente de imagens reais, mas em situações que envolvem a atuação de dois modelos, um de detecção e outro para classificação das células.

Uma possível causa para isso seria o modelo ter sido treinado com essas três classes juntas, sendo elas com lesão, não lesão e não célula. Para uma possível melhora poderiam ser desenvolvidos dois modelos distintos. O primeiro seria responsável pela identificação de células, tendo somente duas classes: célula e não célula. Então, as caixas que esse primeiro modelo classificasse como célula seriam passadas para o segundo modelo, treinado para diferenciar as células com lesão das células sem lesão.

5. Conclusão

Os modelos desenvolvidos variando os diferentes recortes retornaram bons resultados. Para concluir isso podemos comparar já que em [Diniz et al. 2021a] são utilizadas imagens segmentadas para classificação de um sistema de duas classes, tendo o seu melhor modelo atingido 0,96 de acurácia. Já o nosso melhor modelo foi o *EfficientNetB0*, o qual, com tamanho de caixa 70x70, atingiu uma acurácia de 0,94. Dessa maneira, conclui-se que é possível ter métricas com valores adequados sem a necessidade de utilizar informações de segmentação das imagens.

Em relação à investigação do uso dos modelos e os recortes, conseguimos perceber que na maioria dos modelos todos os resultados ficaram bem próximos uns dos outros, sendo que o modelo *EfficientNetB0* alcançou os melhores resultados em dois tipos de tamanho de caixa. Já em relação ao tamanho das caixas, a que proporcionou melhores resultados foi a de tamanho 70x70.

Em especial, os resultados e acurácia nos testes foram acima de 90%, enquanto a precisão foi em torno de 90%, indicando que a abordagem adotada possui bom compor-

tamento com imagens reais de citologia convencional, inclusive incluindo sobreposição e outras estruturas.

Dessa maneira, podemos dizer que nossa proposta pode se caracterizar uma ferramenta útil na análise de células cervicais lesionadas em imagens de Papanicolau na rotina de laboratório. Sua implementação pode auxiliar a tomada de decisão dos profissionais da saúde.

6. Agradecimentos

Agradecemos o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, da UFOP e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Referências

- Diniz, D., Keller, B., Rezende, M., Bianchi, A., Carneiro, C., Oliveira, R., Luz, E., Ushizima, D., Medeiros, F., and Souza, M. (2022). A cytopathologist eye assistant for cell screening. *AppliedMath*, 2:659–674.
- Diniz, D. N., Rezende, M. T., Bianchi, A. G. C., Carneiro, C. M., Luz, E. J. S., Moreira, G. J. P., Ushizima, D. M., de Medeiros, F. N. S., and Souza, M. J. F. (2021a). A deep learning ensemble method to assist cytopathologists in pap test image classification. *Journal of Imaging*, 7(7):111.
- Diniz, D. N., Rezende, M. T., Bianchi, A. G. C., Carneiro, C. M., Ushizima, D. M., de Medeiros, F. N. S., and Souza, M. J. F. (2021b). A hierarchical feature-based methodology to perform cervical cancer classification. *Applied Sciences*, 11(9).
- Gómez, O. H., Sánchez-DelaCruz, E., and de la Mata, A. P. (2017). Classification of cervical cancer using assembled algorithms in microscopic images of papanicolaou. *Res. Comput. Sci.*, 139:125–134.
- Isidoro, D. W. A. (2021). Classificação automática de recortes de células cervicais baseada em características não geométricas. Trabalho de Conclusão de Curso (Ciência da Computação), Departamento de Computação, Universidade Federal de Ouro Preto, Ouro Preto, MG. Disponível em: <https://monografias.ufop.br/handle/35400000/3075>.
- Mariarputham, E. J. and Stephen, A. (2015). Nominated texture based cervical cancer classification. *Comput Math Methods Med*, 2015:586928.
- Pangarkar, M. A. (2022). The bethesda system for reporting cervical cytology. *Cytojournal*, 19:28. PMID: PMC9168399.
- Rezende, M. T., Bianchi, A. G., and Carneiro, C. M. (2021). Cervical cancer: Automation of pap test screening. *Diagnostic Cytopathology*, 49(4):559–574.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3):209–249.
- Who (2022). World Health Organization - Cervical Cancer. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/cervical-cancer>. Acessado em: 20-03-2023.