Abordagem Few-Shot Learning para diagnóstico de câncer de colo do útero em imagens celulares segmentadas

Murilo S. Ortiz [1], Alaine M. Guimarães[1], Andréa T. dos Santos[2], Ednéia P. Machado[2], Daniela Etlinger-Colonelli[3], José Carlos F. da Rocha[1]

[1]Departamento de Informática (DEINFO) - Universidade Estadual de Ponta Grossa (UEPG), Ponta Grossa, Brasil (21009326@uepg.br)
[2]Departmento de Análises Clínicas (DECLIN)- Universidade Estadual de Ponta Grossa (UEPG), Ponta Grossa, Brasil
[3]Centro de Patologia - Instituto Adolfo Lutz, São Paulo, Brasil

Resumo: O câncer de colo do útero acomete milhares de pessoas do sexo feminino todos os anos. O método convencional de detecção, através da coleta e análise de amostras, pelo exame Papanicolau, possui características manuais que podem gerar uma significativa taxa de falsos negativos. Métodos de Aprendizado de Máquina têm sido usados em pesquisas visando automatizar a classificação das células dessas amostras, porém, as bases de imagens, consideradas pequenas, têm se mostrado um fator limitante. Esse trabalho propôs a criação de um modelo computacional de classificação citopatológica baseado na abordagem Few-Shot Learning. A acurácia obtida em torno de 90% está alinhada com resultados esperados na literatura.

Palavras-chave: Few-Shot Learning; ResNet50; VGG19; Câncer de colo de útero.

INTRODUÇÃO

O câncer de colo do útero (ou câncer Cervical) ´e o quarto tipo de câncer que mais acomete pessoas do sexo feminino ao redor do mundo, segundo dados da Agência Internacional de Pesquisa sobre o Câncer referentes ao ano de 2022 (IARC 2024). No Brasil, ´e o terceiro tipo de câncer mais incidente entre a população feminina (INCA, 2024), com cerca de 6,1% do total, na mortalidade proporcional por câncer em mulheres. Para o ano de 2023, foram estimados cerca de 17mil novos casos, segundo dados do Instituto Nacional de Câncer (INCA, 2024).

O diagnóstico do câncer de colo de útero pode ser obtido pelo exame preventivo Papanicolau, um método relativamente barato e eficaz na detecção de lesões precursoras (Sachan et al., 2018), consistindo na coleta e análise visual das amostras celulares do colo uterino baseado no sistema Bethesa de classificação (Nayar e Wilbur, 2015), que visa uniformizar a nomenclatura nacional e internacional, contribuindo para o desempenho dos laboratórios (Brasil, 2012).

Por causa das características visuais, o processo pode ser demorado e desgastante, pois muitas vezes vários profissionais precisam ser consultados para obter-se um diagnóstico preciso, além de que um técnico possui significativas quantidades de lâminas a serem analisadas. Sendo assim, os resultados apresentam muitos falsos negativos, prejudicando a detecção precoce do tumor bem como seu tratamento (Rezende et al., 2021).

O Aprendizado de Máquina (ML - do inglês *Machine Learning*) é a ciência de programar computadores para que eles possam aprender através de dados sem serem explicitamente programados (Géron, 2019), permitindo assim, o reconhecimento de padrões nos dados apresentados. Os avanços na área de ML e o desenvolvimento das Redes Neurais Convolucionais (CNN - do inglês *Convolutional Neural Networks*), que permitem a aplicação do aprendizado em imagens, revolucionaram a área de análises biomédicas. Porém, esses algoritmos normalmente dependem de milhares de imagens classificadas para que possam gerar métodos robustos, uma tarefa árdua e em alguns casos nada prática para os estudos clínicos buscados (Medela et al., 2019).

O diagnóstico de câncer e os métodos de Inteligência Artificial (IA) são áreas de rápida evolução e com importantes fatores de convergência. Muitos estudos indicam que as técnicas de escâner aliadas a métodos de ML aumentam as taxas de diagnóstico rápido (Hunter et al., 2022), porém sendo ainda dependentes da qualidade da segmentação e da eficácia da rotulação das células por especialistas (Liang et al., 2019).

Diversas pesquisas voltadas para a detecção de câncer de colo do útero foram realizados nas últimas décadas obtendo resultados expressivos trabalhando com modelos de ML como redes neurais artificiais tradicionais, K-nearest Neighbors (K vizinhos mais próximos), árvores de decisão, máquina de vetor de suporte e, mais recentemente, o aprendizado profundo (*Deep Learning*) envolvendo diferentes tipos de redes como as redes neurais residuais (ResNet), dentre outras (Allahqoli et al., 2022). Na maioria destas pesquisas o fato de a base de dados ser escassa, por dificuldade de obtenção das imagens referentes às lâminas amostrais, sempre e mostrou como um fator limitante para a obtenção de melhores resultados.

Uma abordagem de ML utilizada quando a base de dados é reduzida é o *Transfer Learning*} (TL), que permite que os modelos sejam treinados com menor custo computacional e maior velocidade. Nessa abordagem, uma rede pré-treinada, em uma base de dados grande e variada, como a *ImageNet*, é utilizada. As imagens são adicionados de forma a terem suas características extraídas e de forma a fazer as camadas finais da rede se adaptarem ao domínio desejado (Weiss et al., 2016)..

Também, métodos de *Few-Shot Learning* (FSL), ou *k-Shot Learning*, onde k especifica o número de pontos de dados em cada uma das classes sendo consideradas, tratam da aprendizagem para classificação de objetos, em situações onde apenas algumas amostras de exemplos estão disponíveis (Wang et al., 2020). Estes métodos são úteis justamente para casos em que a aquisição de uma base de dados em larga escala é difícil ou inviável, como no diagnóstico citopatológico de células do colo uterino.

Uma das técnicas de FSL que pode ser empregada para a classificação de câncer cervical é a das redes Siamesas onde duas ou mais redes convolucionais são utilizadas para a extração das características e uma camada adicional calcula a similaridade entre as entradas através de uma função de perda. Durante o treinamento de uma rede Siamesa, cada rede de extração retorna um conjunto de dados sobre as entradas e o cálculo da perda se dá através desses dados. O objetivo, durante o treinamento, é reduzir ao máximo a perda (Bromley et al., 1993). Em casos onde as bases de dados são desbalanceadas ou pequenas demais é possível utilizar redes de extração pré-treinadas como é o caso das redes VGG19 (Simonyam e Zisserman, 2015) e ResNet50 (He et al., 2016).

Medela et al. (2019) verificaram que, para bases de dados reduzidas, com poucas amostras por classes, algoritmos de redes Siamesas utilizados com redes pré-treinadas como a VGG16 performaram melhor do que os algoritmos de TL.

Diniz e Guimarães (2021) aplicaram as redes Siamesas a uma base de dados referente a cancer de colo do útero, porém, devido a uma restrição da base e recursos computacionais limitados, obtiveram uma acurácia para as classes de cerca de 60% apenas. Outro fator de interferência identificado no trabalho foi o fato de as imagens das lâminas de coleta do exame Papanicolau apresentarem diversas células, muitas vezes pertencentes a classes diferentes. Assim, Assim, os resultados demonstraram que o modelo treinado foi influenciado pelos fatores acima, demonstrando uma baixa capacidade de generalização do modelo criado.

Outro fator der interferência é o fato de as imagens das lâminas de coleta do exame Papanicolau apresentarem diversas células, muitas vezes pertencentes a classes diferentes, o que, após aplicados os métodos de data augmentation, tenha levado a uma falsa classificação das imagens inseridas na entrada do modelo. Assim, os resultados demonstraram que o modelo treinado foi influenciado pelos fatores acima, demonstrando uma baixa generalização alcançada.

O objetivo deste trabalho foi criar um modelo computacional para classificação de amostras citopatológicas do colo do útero, coletadas durante a realização do exame Papanicolau, utilizando a abordagem de Aprendizado de Máquina denominada Few-Shot Learning implementada nas redes Siamesas.

MATERIAL E MÉTODOS

A base de dados utilizada é composta por imagens de amostras citopatológicas de exame Papanicolau que foram coletadas no Instituto Adolfo Lutz de São Paulo em microscópio Olympus BX 40 nos aumentos óticos de 400x e 100x, com o sistema de imagem Opticam LOPT 14003 resolução de 14 megapixels e dimensões 4096px x 3286px.

A fim de evitar o diagnóstico errôneo de cada imagem e melhorar os resultados observados em Diniz e Guimarães (2021), visto que, em um mesmo bloco podem ser encontrados diversas células pertencentes a categorias diferentes (Figura 1), foi necessário isolar cada uma das células para o processamento. As imagens foram demarcadas e depois recortadas utilizando um algoritmo computacional desenvolvido na linguagem Python 3.10 e utilizando a biblioteca OpenCV (Bradski e Kaehler, 2008). Assim, foram definidos 1860 blocos de imagens pertencentes a 11 classes diferentes (Tabela 1), de acordo com o sistema Bethesa de classificação (Nayar e Wilbur, 2015).

Tabela 1. Quantidade final de imagens obtidas de cada classe de célula após o recorte

Classificação de Célula	Quantidade
Parabasal	162
Intermediária	132
Superficial	65
Metaplásica	41
LSIL	24
HSIL	451
Carcinoma Não-Queratinizante	314
ASCUS	13
ASCH	33
Carcinoma	180
Carcinoma Microinvasivo	445
Total	1860

Na Figura 1 pode-se observar um exemplo de como o recorte foi realizado, a figura 1.a representa a imagem original da lâmina e a figuras 1.b as células que foram extraídas da mesma, nota-se a presença de classes diferentes na mesma lâmina.

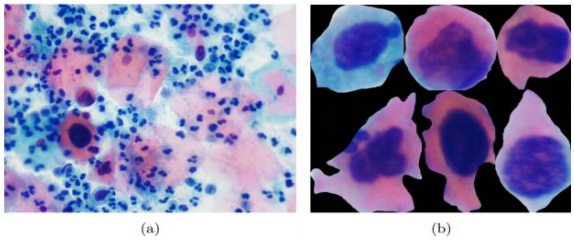


Figura 1. Figura 1. Exemplo do recorte executado na base de dados. Imagem original (a) e os recortes produzidos após o processamento (b). As células possuem as classes (da esquerda pra direita respectivamente), Carcinoma Microinvasivo, HSIL, HSIL (primeira linha), HSIL, HSIL e Carcinoma Microinvasivo (segunda linha).

Tendo em vista que a base de dados se apresentou desbalanceada, com quantidades de imagens por classe variando de 13 (ASCUS) a 451 imagens (HSIL), , foram selecionadas quatro classes a serem consideradas pelos algoritmos de Aprendizado de Máquina, sendo elas: HSIL, Carcinoma Microinvasivo, Carcinoma Queratinizante e Carcinoma Não-Queratinizante. Além disso, as imagens foram rotacionadas em 90º e 180º , que são estratégias utilizadas para aumentar a base de dados existente. As imagens passaram inicialmente por um pré-processamento de adequação ao tamanho 244x224px para a entrada nas redes e uma normalização dos pixels.

Para fins de teste e comparação, foi utilizado um algoritmo padrão de Transfer Learning utilizando como redes de extração a Resnet50 e a VGG19. Era esperado que as redes Siamesas performassem melhor, de acordo com o observado por Medela et al. (2019). A hipótese abordada prevê uma melhora nos resultados de classificação de células cancerígenas, uma vez que, ao separar as células da imagem original, isolando em um bloco de imagem uma instância única das classes, o modelo pode aprender melhor as características de cada classe. Para todos os algoritmos executados a base foi dividida aleatoriamente em 80% para treinamento e validação e 20% para teste.

A arquitetura básica das redes de Transfer Learning e das redes Siamesas foram construídas utilizando como suporte as redes ResNet50 e VGG19 excluindo as camadas finais de classificação. A fim de evitar o overfitting e aproveitar os recursos aprendidos por essas redes, as camadas foram congeladas e a saída passou por uma camada de Pooling2D, a fim de reduzir a dimensão espacial das características extraídas. Também foi aplicada uma camada densa com 512 neurônios e ativação ReLU, para processar as características e uma camada de dropout para regularização. A camada de saída possuía um número de neurônios igual ao número de classes e a ativação softmax foi utilizada para converter as saídas em probabilidades para cada classe. Para as redes Siamesas, o modelo foi compilado com duas redes da mesma arquitetura e as saídas comparadas pela função de perda.

Ambos os modelos utilizaram otimizador de compilação Adam, tamanho de lote 32 e função de perda sparse categorical crossentropy loss, que mede a dissimilaridade entre o modelo calculado e a verdadeira classe das entradas.

Para a análise dos resultados verificou-se o gráfico de acurácia através das épocas, as métricas obtidas no grupo de teste (acurácia, perda e coeficiente Kappa) e a matriz de confusão resultante. Todos os algoritmos foram executados de forma remota no servidor do Laboratório de Computação de Alto Desempenho (LCAD) da Universidade Estadual de Ponta Grossa.

RESULTADOS E DISCUSSÃO

Os algoritmos de TL foram executados sobre a base de imagens recortadas e rotacionadas. As medidas de acurácia e de perda obtidas durante o treinamento e a validação são apresentadas nos gráficos contidos nas figuras 2 (VGG19) e 3 (ResNet50). A base de teste foi executada e os resultados obtidos foram plotados em uma matriz de confusão (figuras 4 e 5).

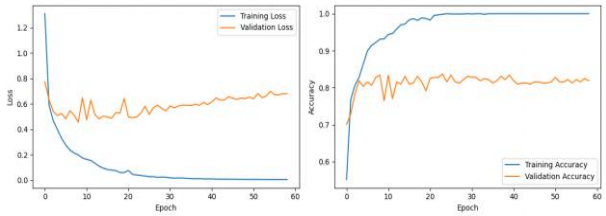


Figura 2. Perda (esquerda) e acurácia (direita) ao longo das épocas para treino e validação da rede Transfer Learning com rede base VGG19.

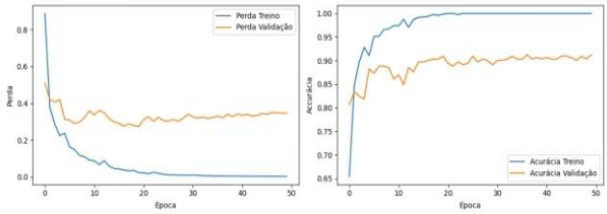


Figura 3. Perda (esquerda) e acurácia (direita) ao longo das épocas para treino e validação da rede Transfer Learning com rede base ResNet50.

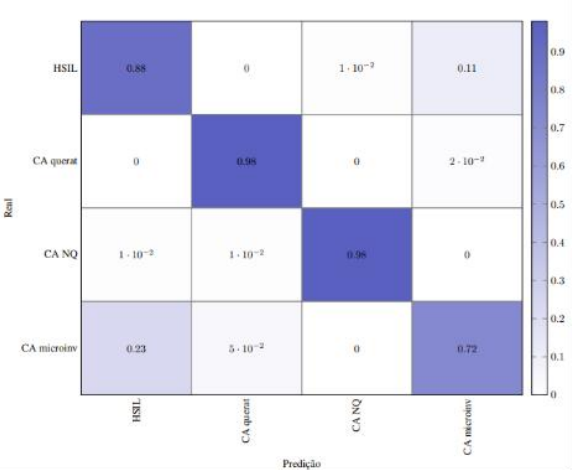


Figura 4. Matriz confusão gerada pelo teste, rede Transfer Learning ResNet50.

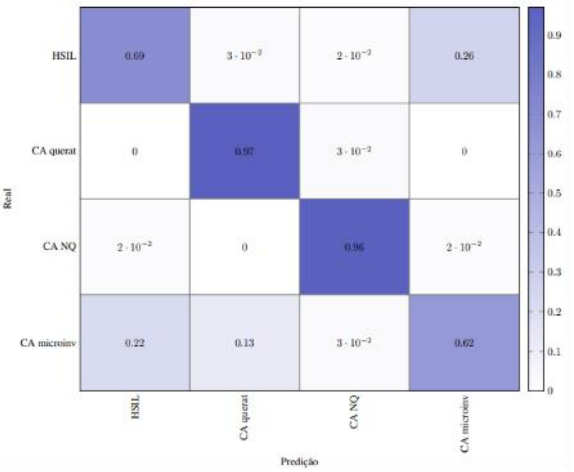


Figura 5. Matriz confusão gerada pelo teste, rede Transfer Learning VGG19.

As métricas obtidas para a execução do TL com VGG19 foram: perda do teste de 63%, acurácia do teste de 82% e coeficiente Kappa de 0,76. Já as métricas obtidas para a execução do TL com ResNet50 foram: perda do teste de 26%, acurácia do teste de 89% e coeficiente Kappa de 0,86.

Pode-se observar para as duas execuções (TL com VGG19 e TL com ResNet50) que as matrizes de confusão marcaram uma confusão maior entre as classes HSIL e Carcinoma Microinvasivo do que para as outras classes. No caso da execução de Transfer Learning com a VGG19, 62% de acerto para Carcinoma Microinvasivo e 69% para a classe HSIL (Figura 4). Já para a execução com a ResNet50 a rede Transfer Learning mostrou uma taxa de acerto de 72% para Carcinoma Microinvasivo e 88% para HSIL (Figura 5).

Para as redes Siamesas os gráficos de aprendizado podem ser visualizados nas figuras 6 e 7 e suas matrizes de confusão nas figuras 8 e 9.

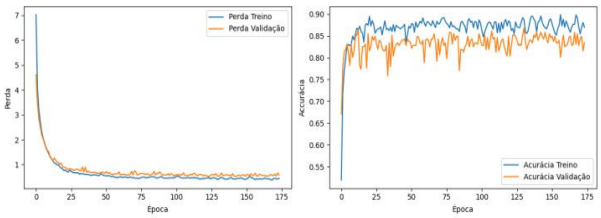


Figura 6. Perda (esquerda) e acurácia (direita) ao longo das épocas para treino e validação da rede Siamesa com rede base VGG19.

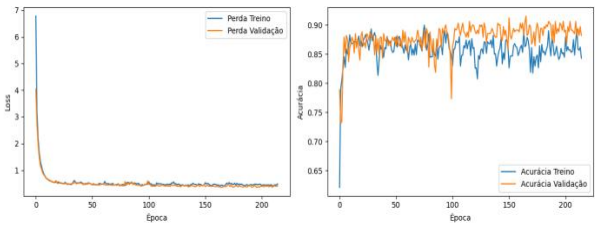


Figura 7. Perda (esquerda) e acurácia (direita) ao longo das épocas para treino e validação da rede Siamesa com rede base ResNet50.

Pode-se observar que as redes Siamesas não alcançaram o overfitting porém baseando-se na perda de validação atingiram o early stopping programado de 50 épocas. As métricas para a rede Siamesa com VGG19 sobre o grupo de teste foram: perda do teste de 58%, acurácia do teste de 84% e o coeficiente Kappa de 0,79. As métricas para a rede Siamesa com ResNet50 sobre o grupo de teste foram: perda do teste de 34%, acurácia do teste de 90% e o coeficiente Kappa de 0,87.

Assim como nos algoritmos de Transfer Learning, as redes Siamesas obtiveram maior confusão entre as classes Carcinoma Microinvasivo e HSIL, sendo 74% de acerto para Carcinoma Microinvasivo e 72%

para HSIL para a execução com a VGG19 e, para a ResNet50, 79% de acerto para HSIL e 86% para Carcinoma Microinvasivo.

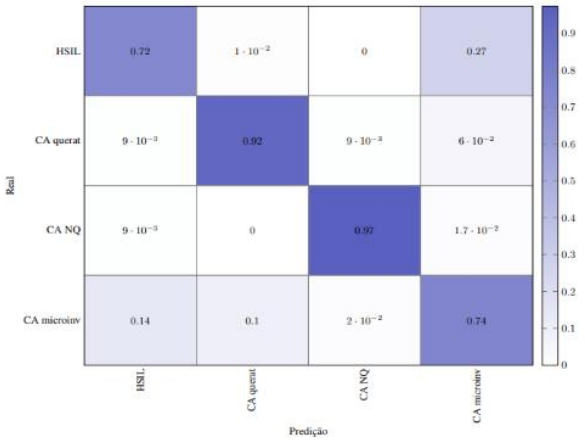


Figura 8. Matriz confusão gerada pelo teste, rede Siamesa VGG19.

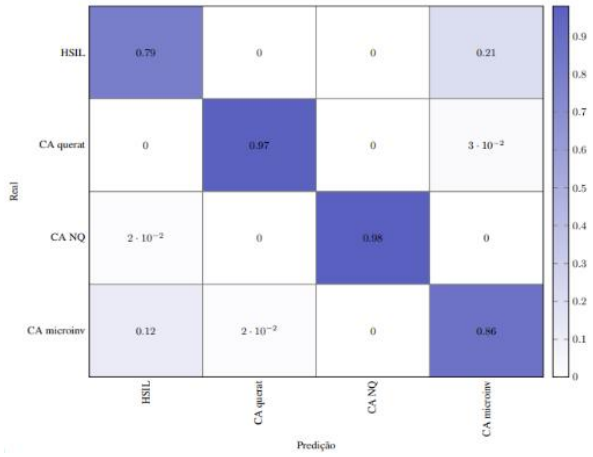


Figura 9. Matriz confusão gerada pelo teste, rede Siamesa ResNet50.

A rede pré-treinada ResNet50 performou melhor que a rede VGG19, por possuir métricas superiores para cada método de ML utilizado. Como encontrado por Medela et al. (2019), as redes Siamesas performaram melhor que as redes de Transfer Learning para as amostras utilizadas. A segmentação (recorte) das imagens influenciou na melhora dos resultados de classificação, em comparação com Diniz e Guimarães (2021), como observado por Liang et al. (2019).

CONCLUSÃO

A partir dos resultados obtidos pelo trabalho, pode-se inferir que o aumento da base de dados e a segmentação (recorte) das imagens permitiu aos modelos treinados uma performance superior à obtida anteriormente por Diniz e Guimarães (2021). Além disso, as métricas da rede ResNet50 foram superiores as da rede VGG19 para o grupo teste analisado. As redes Siamesas obtiveram um resultado levando a uma acurácia de 90%, superior às redes de Transfer Learning, as quais utilizaram a mesma rede pré-treinada como base.

Para trabalhos futuros, sugere-se tratar de diferentes formas o desbalanceamento das classes estudadas e desenvolver um algoritmo que consiga detectar as células carcinomáticas sem a necessidade de marcação e recorte. Além disso, espera-se que os resultados para as classes HSIL e Carcinoma Microinvasivo sejam melhorados significativamente, uma vez que é de interesse a precisadistinção entre as classes correspondents a lesões e a carcinomas.

AGRADECIMENTOS

Agradecemos ao Instituto Adolfo Lutz por ceder as amostras de imagens das amostras citopatológicas, ao Laboratório de Computação de Alto Desempenho da UEPG e à Fundação Araucária pela bolsa de pesquisa concedida.

REFERÊNCIAS

Allahqoli, L., et al.; Diagnosis of Cervical Cancer and Pre-Cancerous Lesions by Artificial Intelligence: A Systematic Review. Diagnostics, Suíça, v. 12,11 2771, 2022.

Bera, K.; Schalper, K. A.; Rimm, D. L., Velcheti, V.; Madabhushi, A.; Artificial intelligence in digital pathology – New tools for diagnosis and precision oncology. Nat. Rev. Clin. Oncol. 2019.

Bradski, G.; Kaehler, A.; Learning OpenCV: Computer vision with the OpenCV library. 1.Ed. Sebastopol: O'Reilly Media, 2008.

Brasil. Inca. Instituto Nacional do Câncer. Nomenclatura brasileira para laudos citopatológicos cervicais. Coordenação Geral de Prevenção e Vigilância, Divisão de Detecção Precoce e Apoio à Organização de Rede. 3. ed., Rio de Janeiro: INCA, 2012.

Bromley, J.; Bentz, J. W.; Lecun, Y., et al.; Signature verification using a siamese time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence, 1993.

Diniz, J. L. M; Guimarães, A. M.; Uso de Few Shot Learning em base de dados tumoral, Trabalho de Conclusão de Curso; (Graduação em Engenharia de Computação), Universidade Estadual de Ponta Grossa, 2021.

Géron, A. Hands-on Machine Learning with Scikit-Learn, Keras, and Tensor Flow. 2.Ed. Sebastopol: O'Reilly Media, 2019.

He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and

Pattern Recognition (CVPR), Las Vegas, Estados Unidos, p. 770–778, 2016.

Hunter, B.; Hindocha, S.; Lee, R.W. The Role of Artificial Intelligence in Early Cancer Diagnosis. Cancers, 2022.

IARC, I. A. F. R. O. C. (2024). Cancer today. Disponível em: https://gco.iarc.fr/today/factsheets-populations. Acesso em: 15 março 2024.

INCA, Instituto Nacional De Câncer José Alencar Gomes Da Silva. Dados e Números sobre Câncer do Colo do Útero – Relatório Anual 2022. Rio de Janeiro: INCA, 2024. Disponível em: https://www.gov.br/inca/pt-br/assuntos/gestor-e-profissionalde-saude/controle-do-cancer-do-colo-do-utero/dados-e-numeros. Acesso em: 14 março 2024.

Liang, Y.; Tang, Z.; Y,An M; Chen, J.; Liu, Q.; Xiang, Y. Comparison-Based Convolutional Neural Networks for Cervical Cell/Clumps Detection in the Limited Data Scenario, Neurocomputing, 2019.

Medela, A. et al. Few Shot Learning in Histopathological Images: Reducing the Need of Labeled Data on Biological Datasets, 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). Veneza, Itália, p. 1860-1864, 2019.

Nayar, R.; Wilbur, D. C. The Bethesda system for reporting cervical cytology: definitions, criteria, and explanatory notes. 3. ed. Springer Cham, Abril 2015.

Rezende, T.; Silva, R.; Bernardo, F.; Tobias, A.H.G.; Oliveira, P.H.C.; Machado, T.M.; Costa, C.S.; Medeiros, F.N.S.; Ushizima, D.M., Carneiro, C.M.; Bianchi, A.G.C. Cric searchable image database as a public platform for conventional pap smear cytology data. Scientific data. 2021, 8:151.

Sachan, P. L.; Singh, M.; Patel, M. L.; Sachan, R. A study on cervical cancer screenings using pap smear test and clinical correlations. Asia Pac J Oncol Nurs. 2018;5(3);337-341.

Simonyan, K; Zisserman, A; Very deep convolutional networks for large-scale image recognition, 3rd International Conference on Learning Representations (ICLR 2015), Computational and Biological Learning Society, p. 1-14, 2015.

Wang, Y. et al. Generalizing from a few examples: A survey on few-shot learning. ACM Computing Surveys (CSUR), v. 53, n. 3, p. 1-34, 2020.

Weiss, K.; Khoshgoftaar, T. M.; Wang, D.; A Survey of transfer Learning, Journal of Big Data, v.3 a.9, 2016.