

Interpretabilidade de modelos de aprendizado de máquina para a classificação de lesões celulares

Pedro A. Euzébio¹, Rafael A. B. de Queiroz¹, Daniela C. Terra¹,
Mariana T. Rezende², Claudia M. Carneiro², Andrea G. Campos Bianchi¹

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

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

pedro.euzebio@aluno.ufop.edu.br, bonfimraf@gmail.com,
daniela.terra@aluno.ufop.edu.br, trevisanrezende@gmail.com,
carneirocm@ufop.edu.br, andrea@ufop.edu.br.

Abstract. *Due to the widespread use of machine learning algorithms to automate decision-making processes, it is important that they are interpretable to prove themselves reliable. However, although learning achieves cutting-edge results in real-world applications, the its excessive number of parameters is not well understood by humans. In this work, the Local Interpretable Model Agnostic Explanations (LIME) method was used for the interpretability of random forest models constructed in the context of cervical cell image classifications. The results obtained were individual explanations of the most representative instances of the database, followed by an analysis of occurrences of the characteristics among the generated explanations.*

Resumo. *Devido à difusão do uso de algoritmos de aprendizado de máquina na automatização de processos de tomada de decisões, torna-se importante que eles sejam interpretáveis para que se mostrem confiáveis. No entanto, embora o aprendizado alcance resultados de ponta em aplicações do mundo real, o eu número excessivo de parâmetros não é bem compreendido pelos humanos. Neste trabalho, empregou-se o método Local Interpretable Model Agnostic Explanations (LIME) para a interpretabilidade de modelos de floresta aleatória construídas no contexto de classificações de imagens de células cervicais. Os resultados obtidos foram explicações individuais das instâncias mais representativas da base de dados, seguido de uma análise de ocorrências das características em meio às explicações geradas.*

1. Introdução

A cada ano, mais de 500.000 mil mulheres são diagnosticadas com câncer do colo do útero e a doença gera mais de 300.000 mortes pelo mundo [WHO 2022, Cohen et al. 2019], sendo a quarta maior causa de mortes relacionadas ao câncer em mulheres. Desse modo, o câncer cervical se mostra um problema de saúde pública extremamente relevante, em que a prevenção por meio de diagnóstico precoce revela-se eficaz.

O advento de exames de rastreamento, tais como o Papanicolau, permitiram o diagnóstico precoce e a diminuição de mortes. Todavia, o processo de análise clínica se mostra árduo devido ao grande número de células presentes em cada amostra [Wright 2007]. Portanto, uma alternativa de apoio ao profissional de saúde é o uso de citologia digital, com o objetivo de tornar o processo de análise de células cervicais mais eficiente e menos custoso.

Vários autores têm se dedicado a utilizar aprendizado de máquina no sentido de propor novas tecnologias para a classificação de células. No entanto, a confiabilidade de um método de aprendizado de máquina é um ponto crucial para a adesão ao modelo [Ribeiro et al. 2016]. A interpretabilidade destes modelos de aprendizado de máquina fornece uma visão de quais características o algoritmo leva em consideração para tomar uma decisão.

A importância da interpretabilidade nos modelos de aprendizado de máquina é destacada especialmente em aplicações de saúde onde as decisões humanas têm um papel fundamental no diagnóstico e no tratamento da doença. Em suma, a interpretabilidade é a capacidade de fornecer explicações de modelos em termos compreensíveis para humanos. Não deve ser confundido com interpretação de algoritmo, que revela como os modelos fazem previsões, como quais atributos são usados para decisões de modelo. Salienta-se que alguns países restringem o uso do aprendizado de máquina devido a muitas preocupações éticas associadas à justiça, prestação de contas e transparência.

No contexto da classificação, as *Convolutional Neural Networks* (CNNs) são amplamente utilizadas nesse cenário, e apresentam o estado da arte. [Chen et al. 2022] fizeram uso das arquiteturas Xception, MobileNet e MobileNetV2 para a classificação de células cervicais usando a base de dados Herlev [Marinakis et al. 2009]. Os modelos de floresta aleatória (em inglês, *Random Forest* (RF)) também possuem aplicações bem sucedidas na medicina. [Guanglu Sun and Lang 2017] fez uso do RF aplicado a 20 atributos extraídos de células cervicais da base de Herlev.

No contexto clínico e de assistência médica, o método de interpretabilidade LIME (*Local Interpretable Model-Agnostic Explanations*) pode desempenhar um papel de apoio aos profissionais de saúde na tomada de decisões médicas. A partir disso, [Barr Kumarakulasinghe et al. 2020] avaliaram como o LIME realiza explicações a partir de um modelo em comparação com explicações feitas por profissionais da saúde de forma independente. Além disso, também foi investigada a relevância clínica e a confiabilidade das explicações obtidas por meio do LIME. Como resultados, houve o indicativo da relevância das explicações, bem como a consonância em relação as explicações feitas por médicos.

Destaca-se que a integração dos especialistas citopatologistas no processo de estruturação das estratégias de interpretação é muito importante, caso contrário, o modelo dificilmente acaba sendo integrado em uma rotina clínica [Vellido 2020]. Logo, algoritmos de classificação celular podem ser submetidos a estratégias de interpretabilidade.

O principal objetivo deste trabalho é empregar uma estratégia de interpretabilidade em modelos de classificação voltados para citologia cervical, em especial em modelos de RF. Para tanto, utiliza-se a classificação hierárquica de núcleos de células cervicais proposta por [Diniz et al. 2021] e o método LIME e extensão como *Submodular Pick*

LIME para gerar interpretações a respeito dos atributos mais significativos para o processo de classificação.

A motivação deste trabalho é ganhar entendimento de quais são as características que mais corroboram para que um classificador de células cervicais baseado em modelos de RF auxiliem na sua tomada de decisão. Assim sendo, não seria preciso fornecer muitas características das células cervicais durante o treinamento do modelo de RF. A hipótese é que o método LIME poderá ajudar nesta tarefa.

O artigo está organizado como segue. Na Seção 2, descreve-se resumidamente o método LIME. A metodologia proposta é apresentada na Seção 3. Os resultados e discussão são delineados na Seção 4. Por fim, as conclusões e trabalhos futuros são apresentados na Seção 5.

2. *Local Interpretable Model-Agnostic Explanations – LIME*

Tendo em vista a concepção de um explicador agnóstico, algumas técnicas visam desempenhar a tarefa de fornecer informações quantitativas e qualitativas de como modelos caixa preta tomam decisões. A partir disso, é notória a percepção da complexidade global destes algoritmos. Sendo assim, muitas vezes sua compreensão é obtida através de um ponto de vista local, gerado pela interpretação de uma instância específica presente na base de dados [Garreau and von Luxburg 2020].

Desse modo, [Ribeiro et al. 2016] desenvolveram o LIME, um método que tem a capacidade de explicar predições de qualquer classificador ou regressor de forma confiável, obtendo o aprendizado de um modelo localmente em torno da predição. Além disso, foi desenvolvida uma extensão do LIME, denominada Submodular Pick LIME (SP-LIME), que consiste em um algoritmo que seleciona um grupo de instâncias representativas com suas respectivas explicações para lidar com o problema de confiabilidade no modelo.

O SP-LIME por sua vez possui como objetivo obter uma compreensão mais ampla do modelo ao explicar um conjunto selecionado de instâncias individuais, visto que o usuário pode não ter disponibilidade de examinar explicações para todas as instâncias da base de dados. A escolha das explicações é feita de modo a cobrir o maior número de atributos importantes, evitando redundâncias que podem ocorrer ao escolher instâncias com explicações similares.

3. Metodologia proposta

Nesta seção é descrita a metodologia adotada neste trabalho. Na Seção 3.1 é descrita a base de dados utilizada e, a Seção 3.2 apresenta como a investigação de interpretabilidade é realizada para cada hierarquia do modelo de classificação.

3.1. Base de Dados: *CRIC Cervix*

A plataforma de banco de imagens CRIC¹, que tem como uma de suas bases de dados o CRIC Cervix, é um consórcio colaborativo entre pesquisadores que visa fornecer coleções de células para a comunidade científica. A coleção de dados CRIC Cervix possui características semelhantes as obtidas nos exames de Papanicolau, imagens com muitas

¹<https://database.cric.com.br>

células e sobreposições provenientes de citologia convencional. Ela contém 400 imagens e 11.534 células classificadas [Rezende et al. 2021] por especialistas.

No que tange à classificação das lesões presentes na base de dados, o sistema de nomenclatura Bethesda [Pangarkar 2022] é utilizado. Sendo assim, as células presentes no CRIC Cervix são classificadas em seis classes, sendo elas: negativo para lesão intraepitelial ou malignidade (NILM); células escamosas atípicas de significado indeterminado, possivelmente não neoplásicas (ASC-US); lesão intraepitelial escamosa de baixo grau (LSIL); células escamosas atípicas, não pode excluir lesão de alto grau (ASC-H); lesão intraepitelial escamosa de alto grau (HSIL); e carcinoma de células escamosas (SCC). A distribuição de amostras por classes conforme fornecido pela base de dados é descrita em [N. Diniz et al. 2021].

Neste trabalho, utiliza-se a base de dados recortada utilizada por [Diniz et al. 2021] e disponível em *CRIC Cervix Cell Classification - Cropped Image*². Esta base de dados foi submetida a diferentes extratores de atributos responsáveis por extrair atributos relacionados a morfologia e a textura do núcleo das células. Os seguintes algoritmos foram empregados: Region Props, descritores de Haralick, Padrões Binários Locais (LBP), Estatísticas de Adjacência de Limiar (TAS), momentos de Zernike, e Matriz de Coocorrência de Nível de Cinza (GLCM), maiores detalhes sobre a extração dos atributos pode ser encontrado em [Diniz et al. 2021].

3.2. Estratégia para análise de interpretabilidade

Como comentado anteriormente, este trabalho usa o resultado de classificação hierárquica de células cervicais apresentado por [Diniz et al. 2021]. O trabalho de [Diniz et al. 2021] investiga quatro métodos tradicionais de classificadores (*Decision Tree*, k-NN, RF e *Ridge*) para núcleos cervicais e relata seus resultados obtidos por uma proposta de classificação hierárquica. A análise estatística realizada mostrou que o algoritmo RF apresentou os melhores resultados para as imagens de núcleos das bases Herlev e CRIC independentemente do número de classes utilizadas.

A Figura 1 apresenta as 4 etapas da classificação de células considerada neste trabalho. As estratégias de interpretabilidade são aqui utilizadas para investigar os resultados de cada classificador da hierarquia, todos implementados por modelos de RF. Na primeira etapa (Etapa 1) as células são separadas entre normais e alteradas. Na segunda etapa (Etapa 2), as alteradas são avaliadas conforme o grau das lesões (baixo e alto grau). Já nas etapas terceira e quarta (Etapas 3 e 4), são discriminadas as classes das lesões. Desse modo, para cada classificador presente nos níveis da hierarquia haverá uma análise de interpretabilidade, totalizando 4 etapas.

Cada etapa do processo de interpretabilidade conta com uma sequência de passos que constroem os resultados, conforme especificado na Figura 2. Desse modo, o processo se repete para cada classificador em 4 passos que são detalhados a seguir.

Passo 1: aplicação do LIME Tabular

Como primeiro passo é necessário fazer uso da função principal que o LIME oferece, chamada de *explainer*, que nos permite chamar uma observação específica e obter uma

²<https://database.cric.com.br/downloads>

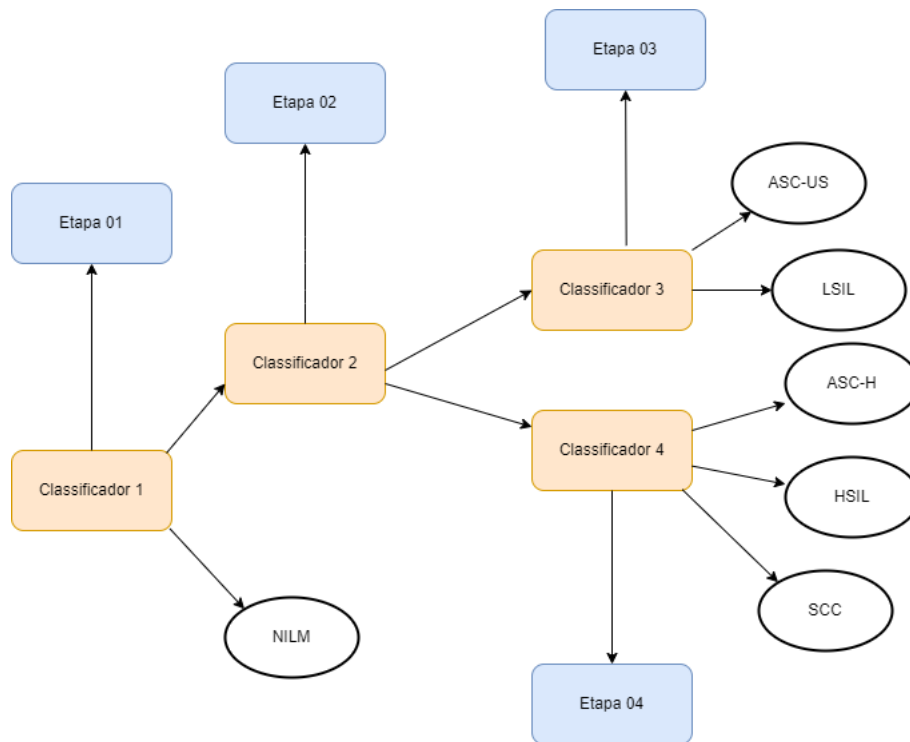


Figura 1. Arquitetura proposta para as explicações.

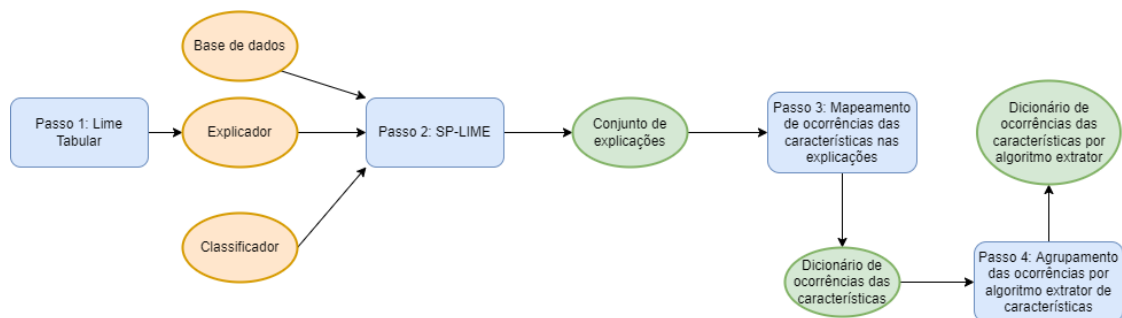


Figura 2. Estrutura de cada etapa dos explicadores apresentados na Figura 1.

interpretação como resultado. Ao instanciar o *explainer* é necessário definir todos os parâmetros manualmente. Para essa aplicação, será inserida a base de dados, além da definição do *mode* do explicador para *classification*, a lista de atributos e a lista de classes. A partir disso, é obtido o explicador que poderá ser usado para prover explicações (vide Figura 2).

Passo 2: aplicação do SP-LIME

A partir do explicador no passo 1, é possível usar o *SP-LIME* por meio da chamada da função principal do módulo *submodular_pick*. Como parâmetros são usados o explicador (modelo RF), a base de dados dos atributos extraídos do núcleo, o método *predict_proba* que retorna a probabilidade da predição para cada classe do modelo, o *num_features* que representa o número de atributos que serão incluídos para a definição das instâncias mais representativas. Além desses, o parâmetro *method* é definido como *full* para indicar que

todas as instâncias da base de dados serão consideradas e o parâmetro *num_exps_desired* para a quantidade de explicações que serão retornadas pelo *SP-LIME*.

Passo 3: Contagem de ocorrências das características contidas nas explicações do SP-LIME

Gerados os conjuntos de explicações, o método *as_map* do LIME foi usado para obter uma lista de tuplas para cada uma das explicações. Cada tupla é formada pelo *id* da característica, como chave, e seu peso como valor. Como foram geradas explicações do SP-LIME retornando apenas a quantidade de características mais importantes de acordo com o valor do atributo *num_features*, foi possível para todas as explicações calcular a ocorrência das características. Desse modo, o total de ocorrências para todas as explicações é dado por: $num_features \times num_exps_desired$.

Como resultado desse passo temos um dicionário em que as chaves são dadas pelo *id* das características que apareceram nas explicações, enquanto os valores são representados pela quantidade de ocorrências do atributo.

Passo 4: Agrupamento das ocorrências por algoritmo extrator de características

Para o último passo, tendo em mãos o algoritmo que gerou a base de dados (tabulares) usada no trabalho de [Diniz et al. 2021], foi possível identificar o *id* das características que faziam parte de cada tipo de algoritmo extrator de atributos. Ou seja, as características extraídas por um mesmo tipo de extrator vão possuir *id* adjacentes indicados explicitamente por um intervalo. Desse modo, o dicionário gerado no passo 3 foi comparado com esse insumo para gerar um novo dicionário, onde as chaves são os nomes dos algoritmos extratores e os valores correspondem a contagens de ocorrências de características de cada um deles.

4. Resultados e discussões

Nesta seção, são apresentados e analisados os resultados da metodologia de explicações para os classificadores aplicados na base *CRIC Cervix*. Como explicado na seção anterior, foi instanciado um explicador pelo LIME Tabular, Passo 1. Como parâmetros foram usados a base de dados *CRIC Cervix* balanceada, com todos os *features* contidos na base para as classes, representadas pelos números de 1 a 6, além do *mode* definido para *classification*. Com o explicador criado, foram geradas as explicações por meio do *SP-LIME* para toda a base de dados (vide Figura 2, Passo 2). Para isso, as explicações foram construídas usando como dados de entrada cada um dos classificadores presentes na hierarquia (Etapas 1, 2, 3 e 4), contando com os recortes das bases de dados utilizadas na classificação.

A partir dessas explicações geradas pelo SP-LIME, o Passo 3 gerou um dicionário com a ocorrência das características. Em cada explicador foram usados todos os atributos (totalizando 232), mas apresentamos nas figuras apenas as 20 características mais presentes no conjunto de explicações devido a limitações de visualização. A seguir, apresentam-se os resultados do Passo 3 obtidos para as diferentes etapas da hierarquia de classificação (Etapas 1, 2, 3 e 4), apresentadas na Figura 1.

Considerando o primeiro classificador da hierarquia, Etapa 1, os resultados obtidos com o explicador para células normais e alteradas são mostrados nas Figura 3.

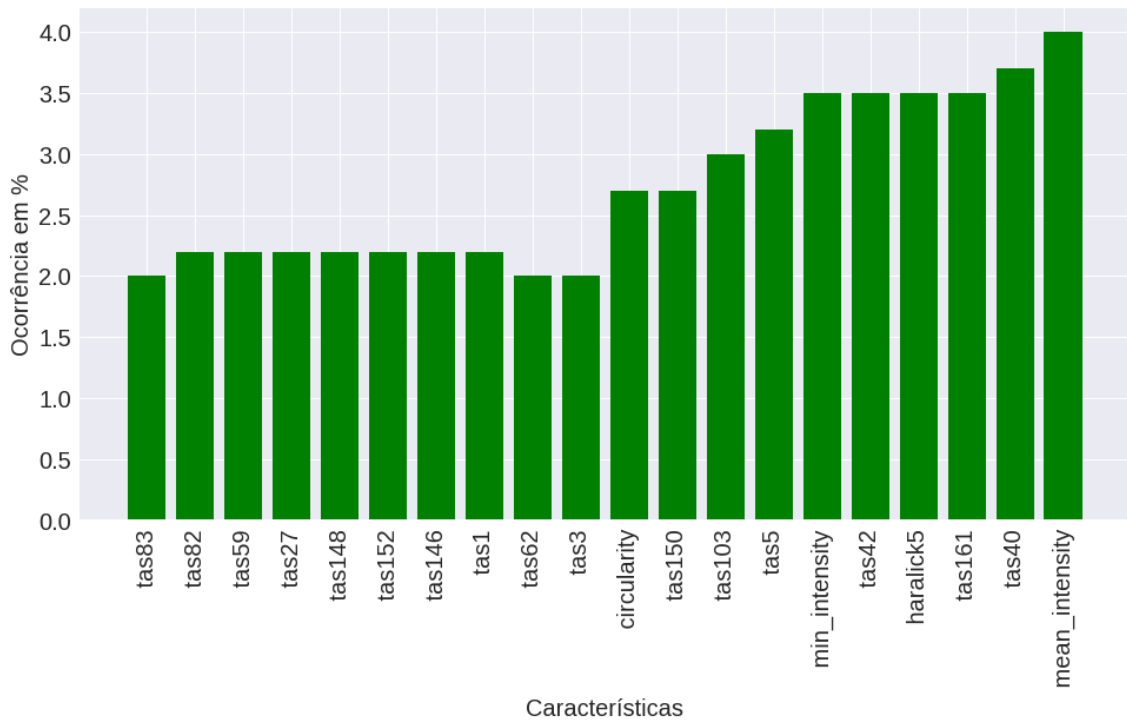


Figura 3. Etapa 1: histograma de características presentes nas explicações do SP-LIME (normais/alteradas).

As classes presentes são a NILM, que representa as células normais, e as alteradas por lesões (ASC-US,LSIL, ASC-H, HSIL e SCC). Nos primeiros 20 atributos mais frequentes notamos que as medidas de TAS representam 2/3 dos dados, seguidos por atributos de intensidade e pela circularidade. Nesse caso, a intensidade média aparece com maior porcentagem de ocorrência.

Na Etapa 2, temos um explicador atuando sobre as células que foram classificadas como alteradas. Nesta fase investigamos um segundo classificador responsável por separar as imagens em células jovens e maduras. As classes de células maduras são ASC-US e LSIL, e, a de células jovens são ASC-H, HSIL e SCC, equivalente a lesões de baixo grau e alto grau, respectivamente. A Figura 4 apresenta os 20 resultados mais frequentes do explicador. Notamos uma equivalência de ocorrência entre as métricas de TAS, LBP e Haralick. Sendo que Haralick possui a maior ocorrência entre todas. Ou seja, a morfologia da célula não é tão importante nesta etapa.

Para as células jovens, Etapa 4, a classificação pode resultar em três classes, tais como: ASC-H, HSIL e SCC. O resultado do explicador pode ser visto em Figura 5. Notamos no gráfico uma equivalência entre as métricas de TAS, Zernike e intensidades. A intensidade média apresenta o maior dentre as 20 medidas.

As células jovens são células alteradas com lesões de alto grau. As classes que representam células jovens são: ASC-H, HSIL e SCC. Seus explicadores são apresentados na Figura 6. Temos um comportamento análogo a Etapa 1 onde das 20 maiores métricas (em ocorrência) 2/3 dos dados estão relacionados à métricas do TAS, seguidos de circularidade e intensidades.

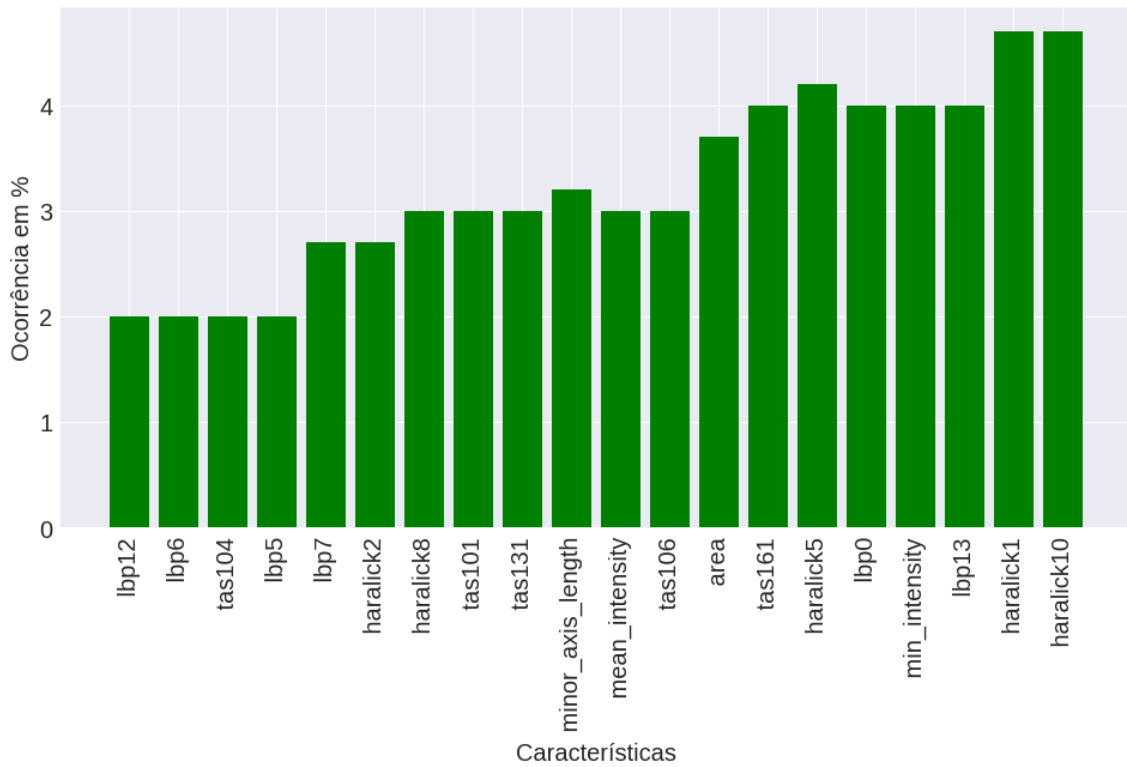


Figura 4. Etapa 2: histograma de características presentes nas explicações do SP-LIME.

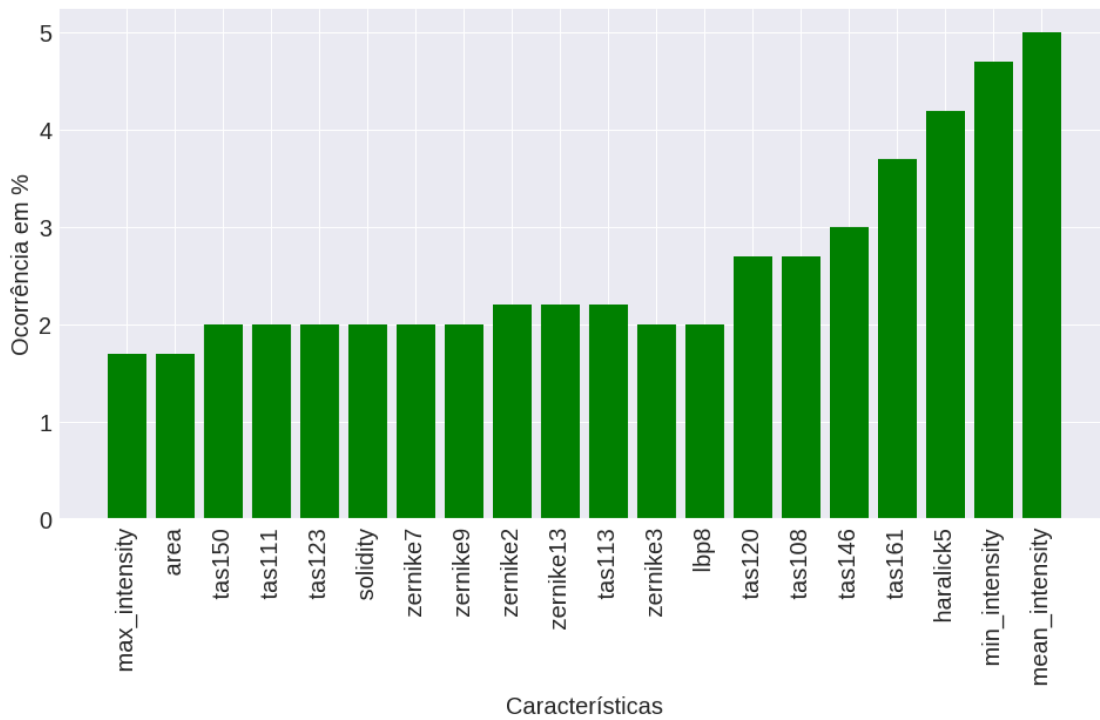


Figura 5. Etapa 3: histograma de características presentes nas explicações do SP-LIME.

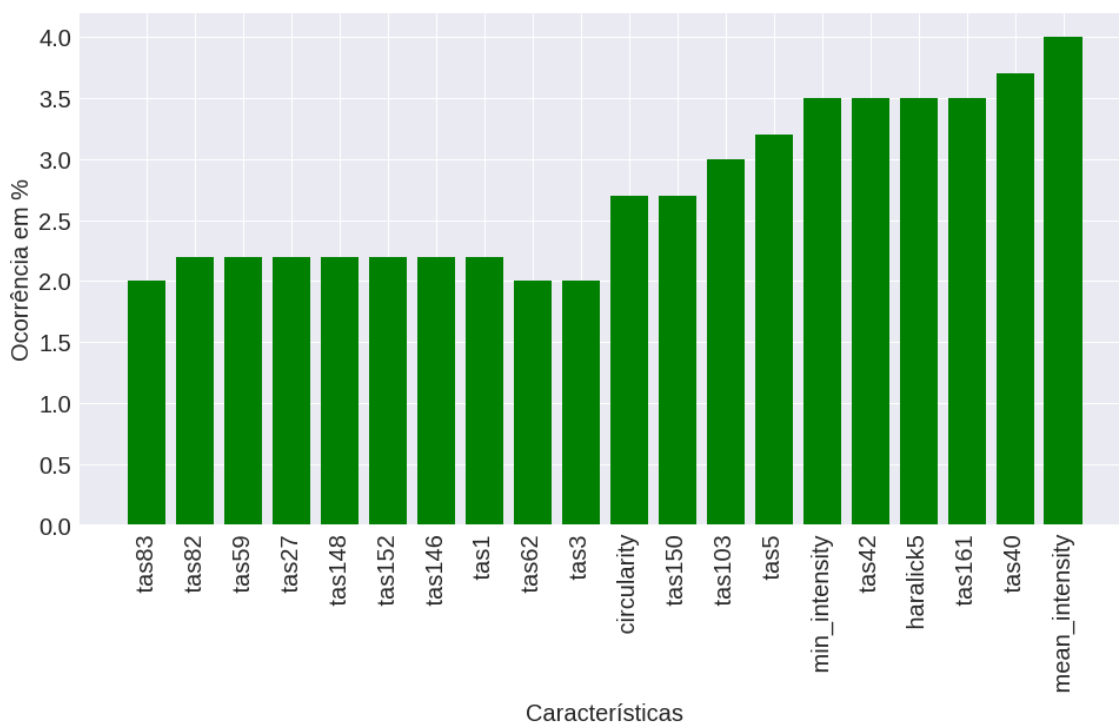


Figura 6. Etapa 4: histograma de características presentes nas explicações do SP-LIME.

Já a partir do Passo 4, criamos gráficos do agrupamento das características que fazem parte de uma mesma família de algoritmos, sendo (*Region Props*, *Haralick*, *TAS*, *LBP*, *Zernike*), ou seja, as informações foram convertidas em histograma, como pode ser visto nos próximos resultados. Os histogramas gerados a partir das explicações do SP-LIME para cada classificador visam entender quais características e quais algoritmos extratores foram mais relevantes no processo de classificação usando RF.

No modelo de floresta aleatória da Etapa 1, onde as células foram diferenciadas entre normais e alteradas houve uma equivalência entre as características morfológicas, as de TAS e LBP. Ressaltadas pela alta ocorrência observada na Figura 7a. Indicando que as variações na morfologia do núcleo e textura são importantes para essa discriminação binária entre lesão e não lesão. Tais informações são evidenciadas pelo que é adotado internacionalmente no sistema Bethesda de classificação de lesões.

Nas três próximas Etapas temos uma predominância dos atributos relacionados ao TAS e uma redução da presença das características morfológicas, ou seja, conforme vamos tornando o modelo mais específico a morfologia do núcleo reduz sua importância e as características relacionadas a sua localização subcelular passam a se destacar na construção do modelo. Essas duas diferenciações acontecem na Etapa 3 e 4, Figuras 5 e 6, respectivamente. Sendo que nas células jovens, as lesões de mais alto grau a diferenciação por meio de métricas de TAS é muito mais significativa para o modelo do que para as lesões de baixo grau.

As métricas de TAS são obtidas pela contagem e estatísticas de pixels acima de um determinado limiar [Hamilton et al. 2007]. Esta métrica têm se mostrado impor-

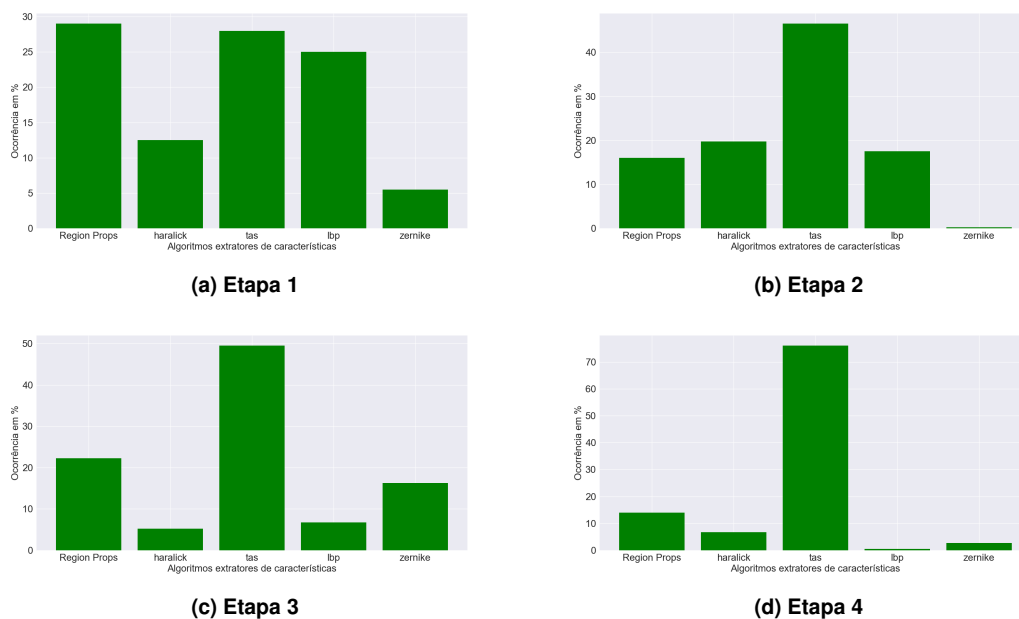


Figura 7. Histograma de extratores de características presentes nas explicações do SP-LIME.

tante para distinguir e classificar imagens que possuam uma textura visual e aparente [Hamilton et al. 2007]. Do ponto de vista do profissional, em lesões de mais alto grau busca-se a hiper Cromasia, que é a situação onde o núcleo se cora mais intensamente devido a maior riqueza de DNA, gerando estruturas de textura mais aparente. Cromatina grosseira e irregular é um dos maiores critérios de malignidade.

Em especial o modelo de floresta aleatória possui em suas explicações regras baseadas nas características. Porém, esses dados poderiam não denotar explicitamente a verdade biológica. Desse modo, para um citopatologista, a interpretabilidade do modelo não é totalmente transparente, visto que seria necessário decodificar os códigos mencionados.

O ponto de vista do profissional em uma rotina de laboratório, a sua interpretação é definida em função do Sistema Bethesda. Este sistema de classificação de lesões celulares é amplamente usado pelos citopatologistas e destaca-se como modelo de laudo padronizado em citologia cervicovaginal. Além de sua reprodutibilidade, reflete a compreensão mais atual do câncer cervical. A característica mais importante é a sua relevância, onde cada classe tem implicações clínicas claras, baseadas em evidências sólidas e consenso mundial.

Destacamos que dentre os critérios adotados pelos profissionais as informações de morfologia celular e hiper Cromasia do núcleo são características que estão em consonância com os resultados encontrados no modelo de RF. Gerando indicações que este modelo possui um comportamento similar ao processo de *screening* de um profissional. Investigações mais profundas sobre o tema e as classificações serão realizadas principalmente na diferenciação de lesões de baixo e alto grau, situação em que a conduta médica varia. Em caso de lesões de baixo grau o exame deve ser repetido em 6 meses, e, em caso de lesões de alto grau o paciente é encaminhado para colposcopia e/ou biópsia. Resultados

imprecisos podem ocasionar demora no início do tratamento e prognóstico desfavorável.

5. Conclusões e trabalhos futuros

Neste trabalho, foi realizada uma análise de interpretabilidade do modelo de classificação hierárquica de células cervicais proposto por [Diniz et al. 2021]. Os resultados do LIME evidenciaram quais características foram mais importantes para os classificadores tomarem as decisões. Além disso, o agrupamento das características em algoritmos extratores forneceu uma visão mais ampla de quais características foram mais relevantes para o modelo, o que pode ajudar a entender qual característica biológica é mais relevante a ser observada para classificação das células cervicais.

A metodologia proposta se mostrou capaz de revelar detalhes sobre o modelo que coincidem com as hipóteses apresentadas pelos profissionais da área de saúde. Uma análise de interpretabilidade foi estruturada em etapas de modo a não ficar restrita apenas a aplicação das ferramentas fornecidas pelo explicador agnóstico LIME.

Como trabalhos futuros, destacam-se: (1) validar as explicações geradas e as ocorrências das características apresentadas neste trabalho com profissionais da saúde, a fim de avaliar biologicamente cada etapa do desenvolvimento; (2) comparar o método LIME com outros explicadores tais como o método Shapley Additive Explanations (SHAP) [Lundberg and Lee 2017], no contexto de citologia cervical; (3) complementar o estudo replicando a metodologia de explicabilidade a propostas de classificação de lesões cervicais baseadas em formatos e contornos celulares, como a abordagem recente de [Terra et al. 2023].

Referências

- Barr Kumarakulasinghe, N., Blomberg, T., Liu, J., Saraiva Leao, A., and Papapetrou, P. (2020). Evaluating local interpretable model-agnostic explanations on clinical machine learning classification models. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 7–12.
- Chen, W., Gao, L., Li, X., and Shen, W. (2022). Lightweight convolutional neural network with knowledge distillation for cervical cells classification. *Biomedical Signal Processing and Control*, 71:103177.
- Cohen, P. A., Jhingran, A., Oaknin, A., and Denny, L. (2019). Cervical cancer. *The Lancet*, 393(10167):169–182.
- 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. (2021). A hierarchical feature-based methodology to perform cervical cancer classification. *Applied Sciences*, 11(9).
- Garreau, D. and von Luxburg, U. (2020). Explaining the explainer: A first theoretical analysis of lime. In Chiappa, S. and Calandra, R., editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 1287–1296. PMLR.
- Guanglu Sun, Shaobo Li, Y. C. and Lang, F. (2017). Cervical cancer diagnosis based on random forest. *International Journal of Performability Engineering*, 13(4):446.

- Hamilton, N., Pantelic, R., Hanson, K., and Teasdale, R. (2007). Fast automated cell phenotype classification. *BMC bioinformatics*, 8:110.
- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc.
- Marinakis, Y., Dounias, G., and Jantzen, J. (2009). Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification. *Computers in Biology and Medicine*, 39(1):69–78.
- N. Diniz, D., T. Rezende, M., G. C. Bianchi, A., M. Carneiro, C., J. S. Luz, E., J. P. Moreira, G., M. Ushizima, D., N. S. de Medeiros, F., and J. F. Souza, M. (2021). A deep learning ensemble method to assist cytopathologists in pap test image classification. *Journal of Imaging*, 7(7).
- Pangarkar, M. A. (2022). The bethesda system for reporting cervical cytology. *Cytojournal*, 19.
- Rezende, M. T., Silva, R., Bernardo, F. d. O., Tobias, A. H., Oliveira, P. H., Machado, T. M., Costa, C. S., Medeiros, F. N., Ushizima, D. M., Carneiro, C. M., et al. (2021). Cric searchable image database as a public platform for conventional pap smear cytology data. *Scientific Data*, 8(1):1–8.
- Ribeiro, M., Singh, S., and Guestrin, C. (2016). “why should I trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 97–101, San Diego, California. Association for Computational Linguistics.
- Terra, D., Lisboa, A., Rezende, M., Carneiro, C., and Bianchi, A. (2023). Shape-based Features Investigation for Preneoplastic Lesions on Cervical Cancer Diagnosis. In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 506–513. SCITEPRESS - Science and Technology Publications.
- Vellido, A. (2020). The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Computing and Applications*, 32:1–15.
- WHO (2022). World health organization - cervical cancer. Accessed = 2022-10-09.
- Wright, T. C. (2007). Cervical cancer screening in the 21st century: is it time to retire the pap smear? *Clinical obstetrics and gynecology*, 50(2):313—323.