

Aprendizado Ativo Aplicado ao Diagnóstico de Lesões Dermatológicas

Lucas G. Batista¹, Fernando F. Cruz¹, Pedro H. Bugatti¹, Priscila T. M. Saito^{1,2}

¹Departamento de Computação,
Universidade Tecnológica Federal do Paraná - Cornélio Procópio (UTFPR)
Avenida Alberto Carazzai, 1640, 86300-000, Cornélio Procópio, PR, Brazil

²Departamento de Computação,
Universidade Federal de São Carlos (UFSCar)
Rodovia Washington Luís, km 235, 13565-905, São Carlos, SP, Brasil

{batistal, fernando.2021}@alunos.utfpr.edu.br,
pbugatti@utfpr.edu.br, priscilasaito@ufscar.br

Abstract. *The present work explores the active learning approach to select more informative samples, reducing the cost for annotation of dermatological lesions. For validation, an extensive experimental evaluation was performed considering different datasets, metrics and strategies for description (using traditional and deep extractors), selection (using criteria based on diversity, uncertainty and representativeness) and classification (using traditional supervised algorithms). Based on the results, it can be concluded that the proposed approach presented a significant contribution by enabling a 65% reduction in the amount of labeled samples required for the learning of the classifier.*

Resumo. *O presente trabalho explora a abordagem de aprendizado ativo, selecionando amostras mais informativas, de forma a reduzir o custo para anotação de lesões dermatológicas. Para validação foi realizada uma avaliação experimental extensiva considerando diferentes conjuntos de dados, métricas e estratégias para descrição (com extratores tradicionais e profundos), seleção (com critérios baseados em diversidade, incerteza e representatividade) e classificação (com algoritmos supervisionados tradicionais). A partir dos resultados, é possível observar que a abordagem proposta apresentou contribuições significativas, possibilitando uma redução de 65% na quantidade de amostras rotuladas requeridas para o aprendizado do classificador.*

1. Introdução

O câncer de pele é um dos mais frequentes no Brasil, correspondendo a aproximadamente 30% de todos os tumores malignos no país. O melanoma, embora represente apenas 3%, é o tipo mais grave, devido à sua elevada possibilidade de provocar metástase (i.e. disseminação do câncer para outros órgãos). O diagnóstico precoce desse tipo de tumor é fundamental para a sobrevivência dos pacientes.

Atualmente, uma das técnicas utilizadas por dermatologistas para diagnosticar lesões da superfície da pele é a dermatoscopia. Tal técnica vem sendo amplamente utilizada, pois permite ao dermatologista realizar o exame de maneira não invasiva e segura,

com uma visualização mais detalhada da lesão em relação ao olho nu. Mesmo com equipamentos que auxiliam os exames, o diagnóstico é totalmente baseado na inspeção visual do especialista, e deve-se levar em consideração a experiência clínica do profissional, pois a interpretação é subjetiva mesmo para dermatologistas experientes. Nesse sentido, nos últimos anos, têm ocorrido um crescente interesse na automatização da classificação de lesões de pele [?, ?, ?, ?, ?, ?]

As abordagens da literatura evoluem desde as mais tradicionais, com o uso de morfologia explorando padrões de cor, textura, tamanho da lesão, entre outros; até as mais recentes envolvendo aprendizado profundo. Dentre as baseadas em aprendizado profundo algumas envolvem propostas de aumento de dados, redes neurais convolucionais próprias ou tradicionais explorando transferência de aprendizado a partir de arquiteturas (tais como: Inception, ResNet, VGG, DenseNet, entre outras). No entanto, para que métodos de aprendizado profundo consigam uma convergência adequada é necessária uma quantidade considerável de amostras anotadas, bem como um considerável custo computacional para o treinamento do modelo.

Abordagens de aprendizado ativo possibilitam a seleção de um pequeno conjunto de amostras mais informativas a serem apresentadas ao especialista para anotação e utilizadas no aprendizado do classificador. Tais abordagens têm sido amplamente utilizadas e bem sucedidas em diversos contextos, permitindo obter acurácias elevadas mais rapidamente e com mínima intervenção do especialista no processo de aprendizado. Apesar de resultados significativos, algumas delas são ineficientes ou até mesmo inviáveis de serem aplicadas, dependendo do contexto e da quantidade de amostras.

Vale ressaltar que, de acordo com o levantamento bibliográfico realizado, foi encontrado apenas um trabalho [?], o qual explora a abordagem de aprendizado ativo para o contexto de lesão de pele. No entanto, tal trabalho aborda apenas a classificação das lesões entre benignas e malignas (melanoma). Não são explorados outros tipos de patologias. Além disso, não realizam uma avaliação experimental extensiva, considerando diferentes estratégias para seleção das amostras mais informativas, bem como melhores descritores e classificadores para cada tipo de lesão.

Nesse contexto, em resumo, as principais **contribuições** do presente trabalho são: i-) proposta de aplicação de abordagem de aprendizado de ativo para o diagnóstico de lesões dermatológicas; ii-) análise de diferentes estratégias de seleção de amostras mais informativas, de forma a obter classificadores mais eficazes e eficientes para o contexto; iii-) análise e obtenção das melhores combinações extrator-classificador que descrevem e classificam cada lesão de pele; iv-) validação da proposta considerando avaliação experimental extensiva utilizando diferentes conjuntos de dados e métricas/cenários.

2. Metodologia Proposta

O presente trabalho explora a abordagem de aprendizado ativo para a classificação de lesões de pele, de forma a selecionar amostras mais informativas para o processo de aprendizado. Além disso, realiza um estudo para obtenção das melhores características que descrevam cada lesão de pele, bem como os classificadores mais adequados para tal contexto. A Figura 1 ilustra as etapas realizadas no presente trabalho.

A partir do conjunto de dados inicial, foi realizada a etapa de pré-processamento. Na etapa de pré-processamento foi realizado o aumento de dados e a segmentação dos

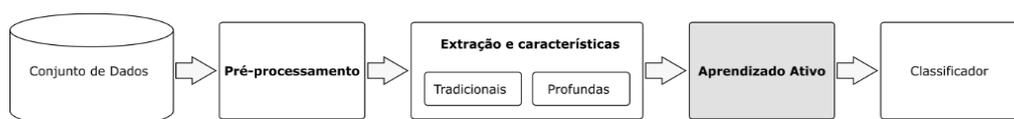


Figura 1. Pipeline da abordagem proposta para classificação de lesões de pele, utilizando técnicas de aprendizado ativo

mesmos, originando os (2) conjuntos de dados utilizados nos experimentos. A descrição dos conjuntos de dados inicial e gerados é apresentada na Seção 3.1.

Após a obtenção dos conjuntos de dados preprocessados, é realizada a etapa de **extração de características**, de forma a obter aquelas que melhor descrevam cada lesão de pele. Para este trabalho foram exploradas propriedades inerentes às imagens, utilizando descritores de baixo nível baseados em cor, textura e forma (i.e., *handcrafted features*), bem como o uso de transferência de aprendizado para arquiteturas CNNs encontradas na literatura treinadas no *ImageNet* para obtenção das *deep features*. A Tabela 1 elenca os descritores utilizados.

A partir dos vetores de características obtidos na etapa de extração, inicialmente, foram avaliados os melhores cenários e configurações por meio da abordagem supervisionada tradicional, de forma a serem considerados posteriormente. Para tanto, além de avaliar quais características (*handcrafted* ou *deep*) melhor descrevem as lesões, foram avaliados quais os classificadores de padrões apresentam melhores resultados, bem como as combinações (extrator-classificador) mostram-se mais apropriadas.

Sendo assim, em seguida, foram exploradas as **abordagens de aprendizado ativo** existentes na literatura (descritas na Seção 3.2) para avaliar a utilização das mesmas no contexto de lesões de pele. Cada estratégia de aprendizado ativo adota critérios para seleção das amostras mais informativas, de forma a acelerar o processo de aprendizado do **classificador** e minimizar a interação do especialista no processo de anotação das amostras a cada iteração do aprendizado.

3. Materiais e Métodos

3.1. Aquisição dos Conjuntos de Dados

Como conjunto de dados inicial foi considerado o *Human Against Machine with 10000 training images* (HAM10000) [?]. É o mesmo conjunto utilizado no evento ISIC 2018 [?]. Possui 10.015 imagens dermatoscópicas coletadas pelo departamento de Dermatologia da Universidade Médica de Vienna, Austria, e da prática clínica do professor Cliff Rosendahl em Queensland, Austrália. Encontra-se dividido em (7) lesões de pele encontradas na prática da dermatologia.

Trata-se de um conjunto de dados desbalanceado em que apenas uma das classes representa mais de 60% de todo o conjunto. Como observado na literatura, o desbalanceamento pode impactar no processo de aprendizado em problemas de classificação. Nesse contexto, foram aplicadas estratégias de **aumento de dados** para as imagens dos tipos *akiec*, *bcc*, *bkl*, *df*, *mel* e *vasc*, com exceção para a categoria *nv*, visto que possui a maior quantidade de amostras.

Para tanto, inicialmente, a partir das amostras originais (conjunto HAM10000),

foram geradas novas amostras por meio da aplicação de operações, tais como redimensionamento na largura e altura; rotações na imagem até 180° ; mudança na orientação horizontal e vertical; aumento na imagem (zoom); dentre outras.

Após a etapa de aumento de dados, foram avaliadas (2) estratégias de **segmentação**, sendo obtidos dois conjuntos de dados D_1 e D_2 , respectivamente. A primeira estratégia de segmentação considera o modelo CNN U-Net proposto por [?]. Para treinamento deste modelo como mecanismo de segmentação semântica, foi utilizado o conjunto de dados do *ISIC Challenge 2017* [?] que possui 1.991 amostras de imagens de lesões de pele anotadas com *ground truth*. A arquitetura requer poucas épocas de treinamento para obter resultados consideráveis, sendo treinada por 30 épocas. Com o modelo treinado, as imagens foram segmentadas por meio de um processo totalmente automatizado, definindo o conjunto D_1 . A segunda estratégia de segmentação utiliza máscaras binárias de segmentação (disponibilizadas por [?]), resultantes da segmentação semântica utilizando FCN e revisão manual do especialista. A Figura 2 apresenta exemplos a partir de cada uma das classes de imagens segmentadas dos conjuntos D_1 (após a obtenção das máscaras binárias resultantes do modelo U-Net.) e D_2 (após a utilização das máscaras disponibilizadas por [?]).

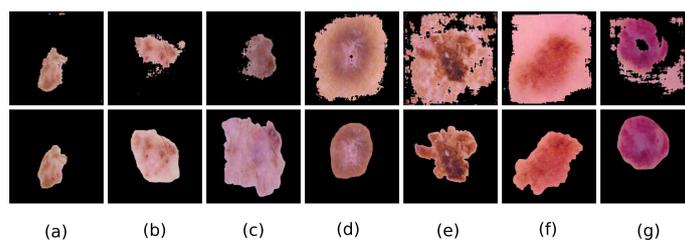


Figura 2. Exemplos de imagens a partir de cada uma das classes: (a) akiec, (b) bcc, (c) bkl, (d) df, (e) mel, (f) nv, (g) vasc dos conjuntos de dados D_1 (imagens superiores) com aplicação de segmentação utilizando a arquitetura U-Net e D_2 (imagens inferiores) com aplicação de segmentação com revisão manual do especialista.

3.2. Aprendizado Ativo

Para validação da abordagem de aprendizado ativo proposta foi realizada uma avaliação experimental, considerando diferentes cenários, além dos conjuntos de dados descritos na Seção 3.1. Para tanto, foram realizados experimentos preliminares, incluindo diferentes extratores de características e a abordagem de aprendizado supervisionado tradicional. Sendo assim, as melhores configurações e combinações (extrator-classificador) são aplicadas nos experimentos envolvendo a abordagem proposta.

Foi adotado o mesmo protocolo experimental, de maneira a realizar comparações justas entre as abordagens de aprendizado. Os resultados foram obtidos a partir de uma média de 5 execuções, a partir dos conjuntos de amostras gerados aleatoriamente, de forma a compor os conjuntos de aprendizado e de teste. Foi realizada a divisão estratificada, mantendo a mesma proporção (80% e 20%) de amostras de cada classe em cada um dos conjuntos (de aprendizado e de teste, respectivamente).

A partir da obtenção das características (por meio dos extratores apresentados na Tabela 1), o primeiro experimento considera a abordagem de aprendizado supervisionado

tradicional, a qual exige o conjunto inteiro anotado. Para tal experimento, foram utilizados os classificadores e suas respectivas parametrizações: k -NN ($k = 5$), RF (criterion = gini e estimators = 100) e SVM (linear). Por meio desse experimento, é possível obter um *baseline* para os demais, indicando qual classificador supervisionado apresenta um melhor desempenho, bem como quais características descrevem melhor os conjuntos.

O segundo experimento explora a abordagem de aprendizado ativo para seleção das amostras mais informativas, com base em diferentes critérios (incerteza, diversidade, representatividade) adotados para a seleção. As estratégias avaliadas foram *Entropy* (EN) [?], *Least Confidence* (LC) [?], *Margin Sampling* (MS) [?], *Root Distance-Based Sampling* (RDS) [?]. Além disso, foi também avaliada a estratégia *Rand*, em que amostras aleatórias são selecionadas a cada iteração. Para cada uma das estratégias são selecionadas por iteração uma quantidade de amostras equivalente a duas vezes a quantidade de classes existentes nos conjuntos de dados (de forma a serem justas as comparações entre as estratégias, bem como de forma a obter amostras representativas de cada classe desde a primeira iteração, conforme sugerido em [?]). Por exemplo, para os conjuntos de dados (D_1 e D_2), os quais apresentam 7 classes, são selecionadas 14 amostras a cada iteração. Tal quantidade de amostras selecionadas a cada iteração é a mesma quantidade de grupos gerados pelo agrupamento considerado por algumas estratégias de aprendizado ativo. Diferentes algoritmos de agrupamento podem ser aplicados. Nos experimentos foi utilizado o algoritmo k -means. Além dos valores de acurácias médias, são consideradas outras métricas para avaliação (e.g. precisão, revocação, F1-Score, quantidades de amostras corrigidas pelo especialista, tempo para seleção das amostras para cada iteração do aprendizado)

As comparações entre as estratégias de aprendizado ativo foram realizadas considerando triplas separadas por hífen, indicando as técnicas utilizadas em cada etapa de extração, seleção e classificação. Por exemplo, a tripla BIC-MS-RF, indica que foram considerados o extrator BIC, a estratégia de aprendizado ativo MS e o classificador RF.

4. Resultados

Com relação ao primeiro experimento realizado, envolvendo a abordagem de aprendizado tradicional, observa-se que os extratores tradicionais, mais especificamente o extrator BIC apresentou um melhor desempenho para ambos os conjuntos D_1 e D_2 (Tabela 1). É possível também observar que, de forma geral, o classificador RF apresentou os melhores resultados de acurácias em comparação aos demais (k -NN e SVM) para ambos os conjuntos de dados, considerando os extratores tradicionais. Percebe-se que o classificador k -NN apresentou maiores acurácias, em relação ao RF, considerando as características obtidas por meio das arquiteturas profundas. Realizando comparações de desempenho entre os conjuntos de dados D_1 e D_2 (Tabela 1), pode-se observar que o conjunto D_2 apresentou um melhor desempenho. Vale destacar que, em tal conjunto, as imagens foram inicialmente segmentadas automaticamente e posteriormente receberam avaliação e correção manual do especialista.

A Tabela 2 apresenta outras métricas de avaliação além da acurácia, tais como precisão, sensibilidade e F1-Score obtidas pelas melhores configurações (i.e. combinação extrator-classificador) para cada conjunto de dados. É possível observar que o conjunto D_2 (formado por imagens segmentadas automaticamente e com correções manu-

Tabela 1. Resultados considerando a abordagem de aprendizado supervisionado tradicional. Dimensionalidade e média de acurácias \pm desvio padrão obtidos pelos classificadores (k -NN, RF e SVM), utilizando as características obtidas pelos extratores tradicionais e arquiteturas profundas, para os conjuntos de dados D_1 e D_2 . Os melhores resultados (i.e. classificadores) para cada extrator são destacados em negrito. Os melhores resultados (i.e. extratores) para cada classificador são apresentados sublinhados. Os melhores resultados (i.e. maiores acurácias) obtidos são apresentados com asterisco

Extrator	Dimensão	D_1			D_2		
		RF	k -NN	SVM	RF	k -NN	SVM
ACC	768	84.03\pm0.64	74.84 \pm 10.46	42.76 \pm 1.24	86.98\pm0.24	83.68 \pm 0.50	45.27 \pm 0.56
BIC	128	87.43\pm0.42*	75.64 \pm 0.62	44.14 \pm 0.62	92.13\pm0.24*	82.02 \pm 0.44	46.29 \pm 0.60
CEDD	144	<u>64.63\pm0.66</u>	64.38\pm0.52	49.98 \pm 0.46	71.51\pm0.78	69.54 \pm 1.02	55.49 \pm 0.52
FCTh	192	67.34\pm0.66	64.53 \pm 0.56	45.99 \pm 0.74	77.67\pm0.98	74.24 \pm 1.26	52.38 \pm 0.60
GCH	66	84.13\pm0.68	76.95 \pm 0.42	40.53 \pm 0.98	90.32\pm0.50	84.40 \pm 0.70	41.06 \pm 0.30
HaralickColor	15	69.00\pm0.80	45.62 \pm 0.56	34.62 \pm 0.94	82.39\pm0.34	80.56 \pm 0.58	41.97 \pm 0.66
JCD	168	71.34\pm0.98	69.56 \pm 0.70	50.66 \pm 0.64	79.54\pm0.92	77.23 \pm 0.76	57.28 \pm 0.46
LBP	256	46.95\pm0.58	43.67 \pm 0.78	25.96 \pm 0.64	40.98\pm0.64	36.22 \pm 0.78	23.30 \pm 2.18
LCH	264	67.04\pm0.80	48.07 \pm 0.82	41.38 \pm 1.14	69.14\pm0.82	51.64 \pm 0.60	41.16 \pm 0.52
Moments	4	33.64\pm0.92	28.84 \pm 0.94	24.62 \pm 0.30	47.50\pm0.68	40.28 \pm 0.96	25.10 \pm 0.50
MPOC	18	69.63\pm0.74	45.57 \pm 0.62	29.00 \pm 0.86	82.70\pm0.56	72.61 \pm 0.76	39.91 \pm 0.72
DenseNet121	1024	66.72 \pm 0.92	75.83\pm0.76	67.10 \pm 0.66	79.37 \pm 0.72	90.14\pm0.34	74.42 \pm 1.04
InceptionV3	2048	54.04 \pm 0.66	63.59\pm0.66	59.95 \pm 1.42	65.27 \pm 1.22	81.72\pm0.22	69.02 \pm 0.90
MobileNet	1024	58.43 \pm 0.78	73.15\pm0.48	64.67 \pm 1.10	70.17 \pm 0.98	86.50\pm0.60	71.56 \pm 0.70
ResNet50	2048	64.68 \pm 1.18	74.25\pm0.66	67.06 \pm 0.78	78.64 \pm 1.08	88.54\pm0.68	<u>77.16\pm0.64</u>
VGG16	512	60.80 \pm 0.78	69.28\pm0.76	62.98 \pm 0.74	73.81 \pm 1.20	86.16\pm0.52	69.26 \pm 1.48
VGG19	512	59.72 \pm 1.00	67.85\pm0.88	62.51 \pm 0.56	72.43 \pm 0.76	84.27\pm0.42	68.12 \pm 0.98
Xception	2048	55.02 \pm 1.26	62.75\pm1.00	61.81 \pm 1.22	68.99 \pm 0.44	82.54\pm0.58	70.91 \pm 0.82

Tabela 2. Resultados considerando a abordagem de aprendizado supervisionado tradicional. Métricas de avaliação (acurácia, precisão, sensibilidade, F1-Score) obtidas pelas melhores configurações (i.e. combinação extrator-classificador) para cada conjunto de dados (D_1 e D_2). Os melhores resultados obtidos são destacados em negrito

Conjuntos	Extrator	Classificador	Acurácia	Precisão	Sensibilidade	F1-Score
D_1	BIC	RF	87.43 \pm 0.42	87.96	87.74	87.78
D_2	BIC	RF	92.13\pm0.24	92.63	92.50	92.56

ais realizadas pelo especialista) apresentou os melhores resultados considerando todas as métricas. Sendo assim, a abordagem de aprendizado ativo proposta deve ser avaliada utilizando tal conjunto.

Os resultados apresentados anteriormente foram obtidos por meio de experimentos que consideraram todas as amostras (100%) do conjunto de dados para treinamento dos classificadores. Com outra perspectiva, a abordagem de aprendizado ativo tem como base selecionar amostras mais informativas, com o objetivo de obter um classificador que atinja maiores acurácias com uma menor quantidade de amostras rotuladas, de forma a acelerar o processo de aprendizado e minimizar os esforços e as interações com o especialista no processo de anotação (correção) das amostras.

Considerando os experimentos relacionados à abordagem de aprendizado ativo (Figura 3), para avaliação das estratégias de seleção (EN, LC, MS, Rand e RDS), fixou-se o extrator BIC e o classificador RF dado que os mesmos apresentaram um melhor desempenho no experimento anterior. Dentre as estratégias de aprendizado ativo, MS apresentou as melhores acurácias de classificação.

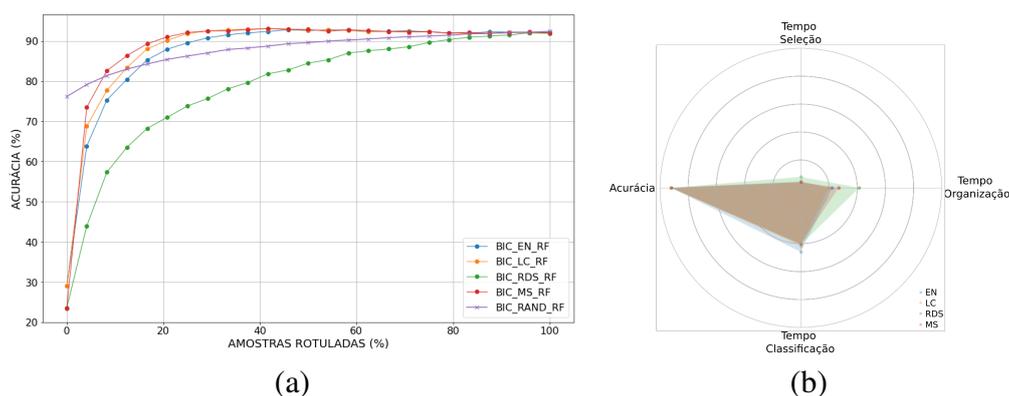


Figura 3. Resultados considerando a abordagem de aprendizado ativo obtidos pelas estratégias de seleção EN, LC, MS, RAND e RDS com a configuração extrator-classificador BIC-RF para o conjunto de dados D_2 . (a) acurácias a cada porcentagem de amostras rotuladas e (b) acurácias e tempos de organização, seleção e classificação de cada estratégia

É possível também observar que as estratégias de seleção (EN, LC, MS e RDS) possibilitam atingir elevadas acurácias mais rapidamente (menos iterações de aprendizado) e com uma menor quantidade de amostras rotuladas (Figura 3(a)), em comparação com a abordagem de aprendizado supervisionado tradicional. Tal abordagem requer uma quantidade maior (100%) de amostras rotuladas para atingir acurácias equivalentes. Por exemplo, para o conjunto D_2 (Figura 3(a)), a estratégia de seleção MS possibilitou uma redução de 65% na quantidade de amostras rotuladas necessárias para o treinamento do classificador, em comparação com a abordagem de aprendizado supervisionado tradicional. A estratégia MS alcançou uma acurácia de 93.24% com apenas 35% (11.494 amostras) do conjunto de treinamento. Enquanto que, na abordagem supervisionada tradicional, o classificador atingiu acurácias de 92.13% sendo necessário 100% (32.242 amostras) do conjunto de treinamento.

Além das acurácias, outras métricas foram consideradas para avaliar o desempenho das estratégias de seleção, a cada porcentagem de amostras, tais como tempos para seleção, organização e classificação (Figura 3(b)). Percebe-se que, em geral, as estratégias (EN, LC e MS) requerem menos tempos para seleção das amostras a cada iteração do aprendizado, por exemplo, em média 0.27, 0.25 e 0.24 segundos, respectivamente. Diferentemente, a estratégia RDS necessita de tempos maiores para seleção, uma vez que realiza o pré-processamento, envolvendo os processos de redução e de organização das amostras baseadas em grafos.

5. Conclusões

O presente trabalho apresenta contribuições significativas para o contexto de lesão de pele. Inicialmente, explorou a abordagem de aprendizado supervisionada tradicional, avaliando diferentes cenários e configurações de extratores (tradicionais e profundos) e classificadores. Foi possível identificar, de forma geral, que o extrator BIC obteve os melhores resultados em extrair padrões das imagens de lesão de pele, nos conjuntos de dados avaliados, juntamente com o classificador RF. Na sequência, observou-se que a abordagem de aprendizado ativo se destaca em alcançar resultados equivalentes e/ou superiores

com uma menor quantidade de amostras de treinamento anotadas. Para trabalhos futuros pretende-se desenvolver novas técnicas de aprendizado ativo, explorando outros critérios de seleção, de forma a selecionar amostras mais informativas.

6. Agradecimentos

O presente trabalho teve o apoio da NVIDIA, CNPq, CAPES, Fundação Araucária, SETI, UTFPR, UFSCar.

Referências

- Adegun, A. and Viriri, S. (2021). Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. *Artificial Intelligence Review*, 54(2):811–841.
- Ali, A.-R. H., Li, J., and Yang, G. (2020). Automating the abcd rule for melanoma detection: a survey. *IEEE Access*, 8:83333–83346.
- Benyahia, S., Meftah, B., and Lézoray, O. (2022). Multi-features extraction based on deep learning for skin lesion classification. *Tissue and Cell*, 74:101701.
- Codella, N. C. F., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., and et al. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*.
- Kaur, R., GholamHosseini, H., Sinha, R., and Lindén, M. (2022). Melanoma classification using a novel deep convolutional neural network with dermoscopic images. *Sensors*, 22(3):1134.
- Popescu, D., El-Khatib, M., El-Khatib, H., and Ichim, L. (2022). New trends in melanoma detection using neural networks: A systematic review. *Sensors*, 22(2):496.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597.
- Saito, P. T., Suzuki, C. T., Gomes, J. F., de Rezende, P. J., and Falcão, A. X. (2015). Robust active learning for the diagnosis of parasites. *Pattern Recognition*, 48(11):3572–3583.
- Settles, B. (2012). *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers.
- Shi, X., Dou, Q., Xue, C., Qin, J., Chen, H., and Heng, P.-A. (2019). An active learning approach for reducing annotation cost in skin lesion analysis.
- Tschandl, P., Rosendahl, C., and Kittler, H. (2018). The HAM10000 dataset: A large collection of multi-source dermoscopic images of common pigmented skin lesions. *CoRR*, abs/1803.10417.
- Tschandl, P., Sinz, C., and Kittler, H. (2019). Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation. *Computers in Biology and Medicine*, 104:111–116.
- Valle, E., Fornaciali, M., Menegola, A., Tavares, J., Bittencourt, F. V., Li, L. T., and Avila, S. (2020). Data, depth, and design: Learning reliable models for skin lesion analysis. *Neurocomputing*, 383:303–313.