CBIS'24
XX Congresso Brasileiro de Informática em Saúde
08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil

Avaliação de variações da rede profunda EfficientNet em bases dermoscópicas

Evaluation of EfficientNet deep network settings on dermoscopic datasets

Evaluación de variaciones de la red EfficientNet en conjuntos dermatoscópicos

Newton Spolaôr[1], Huei Diana Lee[2], Weber Shoity Resende Takaki[1], Claudio Saddy Rodrigues Coy[3], Feng Chung Wu[2]

1 Doutor, Laboratório de Bioinformática, Universidade Estadual do Oeste do Paraná – LABI/UNIOESTE, Foz do Iguaçu (PR), Brasil.
2 Professor Associado-III Doutor, LABI/UNIOESTE, Foz do Iguaçu (PR), Brasil.
3 Professor Titular Doutor, Faculdade de Ciências Médicas, Universidade Estadual de Campinas – FCM/UNICAMP, Campinas (SP), Brasil.

Autor correspondente: (Profa. Dra.) Huei Diana Lee
E-mail: huei.lee@unioeste.br

Resumo

Objetivo: Investigar configurações inéditas da rede profunda EfficientNet-B2 para a classificação de pequenas bases dermoscópicas. **Método**: Uma abordagem para (1) pré-processamento de imagens, (2) classificação com oito configurações para ajuste fino de uma EfficientNet-B2 pré-treinada, e (3) avaliação de classificadores com validação cruzada estratificada em três bases dermoscópicas. **Resultados**: Todos os modelos superaram uma referência experimental, e algumas diferenças estatísticas entre eles foram encontradas. A melhor rede obteve acurácia média de 98,33% no conjunto público PH2. **Conclusão**: Algumas variações inéditas da rede profunda foram consideradas competitivas perante referências recentes em classificação de dermoscopias.

Descritores: Neoplasias Cutâneas; Informática Médica; Inteligência Artificial

Abstract

Objective: To investigate pioneer settings of the EfficientNet-B2 deep network to classify small dermoscopic databases. **Method**: An approach for (1) image pre-processing, (2)

classification with eight settings to fine-tune a pretrained EfficientNet-B2, and (3) classifier evaluation with stratified cross-validation in three dermoscopic databases. **Results**: All the models outperformed a baseline. Some statistical differences among them were found. The best network reached an average Accuracy of 98.33% in the PH2 public dataset. **Conclusion**: Some pioneer configurations of the deep network were found to be competitive against recent references in dermoscopy classification.

Keywords: Skin Neoplasms; Medical Informatics; Artificial Intelligence

Resumen

Objetivo: Investigar configuraciones pioneras de la red profunda EfficientNet-B2 para clasificar pequeñas bases de datos dermatoscópicas. **Método**: Un enfoque para (1) preprocesamiento de imágenes, (2) clasificación con ocho configuraciones para ajustar una EfficientNet-B2 previamente entrenada y (3) evaluación de clasificadores con validación cruzada estratificada en tres bases de datos dermatoscópicas. **Resultados**: Todos los modelos superaron la línea de base. Se encontraron algunas diferencias estadísticas entre ellos. La mejor red alcanzó una precisión promedia de 98,33% en el conjunto público PH2. **Conclusión**: Algunas configuraciones pioneras de la red profunda fueron competitivas frente a referencias recientes en la clasificación de dermatoscopias.

Descriptores: Neoplasias Cutáneas; Informática Médica; Inteligencia Artificial

Introdução

Soluções computacionais com aprendizado de máquina que apoiam a tomada de decisão sobre exames de imagem têm se tornado competitivas. A dermoscopia é um desses exames, sendo útil para auxiliar no diagnóstico precoce de doenças de pele preocupantes e com alta incidência no Brasil, como o melanoma e o carcinoma [1-7].

Parte dessas soluções classificam dermoscopias, diferenciando entre imagens com lesão Maligna (M) e com lesão Não maligna (N). Uma das abordagens mais frequentes nesse contexto tem sido o Aprendizado Profundo (AP). Com AP é possível,

por exemplo, transferir conhecimento adquirido por uma Rede Neural Profunda (RNP) em um vasto repositório, como ImageNet, para enriquecer o treinamento desse classificador sobre um pequeno conjunto de imagens [2,8]. O estudo de AP em conjuntos com poucas imagens é relevante em cenários práticos, como os que envolvem instituições de saúde que (1) possuem um escasso número de imagens adquiridas com o mesmo equipamento e ambiente, e ou que (2) examinam a pele de populações sub-representadas em coleções *benchmark* específicas da área de Dermatologia.

O ajuste fino (*fine-tuning*) de uma rede treinada previamente em um grande repositório para adaptá-la para rotular uma base pequena pode ser útil no aprendizado por transferência. Embora esse ajuste possa ser conectado ao aprendizado *few-shot* [9], *fine-tuning* nem sempre inclui estratégias comuns no último, como o meta-aprendizado.

EfficientNet é uma família de arquiteturas de AP (1) com compromissos variados entre desempenho no ImageNet e custo computacional [10], (2) promissora em domínios como Dermatologia [11-13], e (3) implementada em ferramentas relevantes e gratuitas [14].

O objetivo deste trabalho consiste em investigar configurações inéditas da RNP EfficientNet-B2 para a classificação de pequenas bases dermoscópicas. Parâmetros como o escopo de adaptação no ajuste fino foram variados em oito configurações. Em uma avaliação experimental em bases com 104 a 265 imagens, observaram-se variações competitivas perante referências recentes. Também foi identificado que a melhor rede aprendeu com algumas regiões potencialmente importantes de imagens com lesões.

Métodos

Nesta seção são descritos os materiais usados com a abordagem desenvolvida, além do método em si e dos procedimentos realizados em sua avaliação experimental.

Material

Na Tabela 1 são exibidas propriedades das bases empregadas, incluindo o Complemento do Erro da Classe Majoritária (CECM), obtido ao subtrair de 1 o erro de um classificador ingênuo que rotula imagens com a classe mais frequente. D104 inclui 104 dermoscopias do repositório Derm101 (http://www.derm101.com), já avaliado e

descrito [2,15-18]. PH2 (https://www.fc.up.pt/addi/ph2%20database.html) também já foi investigado [2,6,19]. *The International Skin Imaging Collaboration* (ISIC) (https://www.isic-archive.com) agrupa imagens extraídas de um desafio realizado em 2018 [5,7,20]. Não há necessidade de aprovação de comitê de ética neste trabalho porque todas as imagens dos três conjuntos são de acesso público e não permitem identificação de pessoas.

Tabela 1 – Três conjuntos de dados com números (#) de imagens em cada classe e ao todo

Nome	# imagens malignas	# imagens não malignas	# imagens	CECM
D104	58	46	104	0,5577
PH2	40	160	200	0,8000
ISIC	45	220	265	0,8302

A RNP adotada, a avaliação experimental e as demais implementações são concretizadas no GraphPad Prism 5, Keras 2.13, TensorFlow 2.14 [14], Conda 23.11 (https://docs.conda.io/projects/conda/en/stable), Windows Subsystem for Linux 2, Python 3.10.13 (https://www.python.org) e R 4.3.2 (https://cran.r-project.org/bin/windows/base).

Método Desenvolvido

O método que fundamenta todas as configurações propostas consiste em três etapas: (1) pré-processamento, (2) classificação e (3) avaliação de classificadores.

Na Etapa 1, imagens de um conjunto são convertidas para lotes de vetores multidimensionais normalizados de ponto flutuante. Cada imagem é dimensionada para o formato da entrada da EfficientNet-B2 [10] pré-treinada no ImageNet: 260x260 pixels.

A classificação de lotes por EfficientNet-B2, com funções de ativação *Rectified Linear Unit* (ReLu) e sigmoide [8], é realizada na Etapa 2. Nessa rede existem nove blocos de camadas [10]: um de entrada, sete com estrutura *inverted residual* [21] e um de saída. A base convolucional é constituída pelos Blocos 1 a 8, além das três camadas iniciais do Bloco 9. O processo de ajuste fino da rede pré-treinada consiste em seis procedimentos:

1. Remover as três últimas das seis camadas existentes no Bloco 9, pois elas foram projetadas para rotular imagens com as 1000 classes da ImageNet.
2. Inserir sobre o topo da rede três novas camadas, as quais são estruturadas como as removidas e definem uma rede personalizada na saída do modelo.
3. Congelar os pesos das camadas da base convolucional.
4. Treinar a rede personalizada, com adaptação dos seus pesos e um valor para a Taxa de Aprendizado (TA), em imagens de um conjunto pequeno.
5. Descongelar do início do Bloco *B* até a última camada da base convolucional, após definir *B* conforme a configuração aplicada à RNP.
6. Treinar a rede personalizada junto com as camadas descongeladas, usando um valor particular de TA, em imagens do conjunto pequeno.

Na Tabela 2 são introduzidas as oito configurações propostas neste trabalho para ajuste fino da RNP EfficientNet-B2, as quais representam combinações únicas (1) do índice de bloco *B* e (2) da TA usada no quarto e no sexto procedimentos do ajuste.

Tabela 2 – Oito configurações de rede profunda propostas e investigadas neste trabalho

Nome	*B*=6	*B*=7	*B*=8	*B*=9	TA igual nos procedimentos
FTE1	X				X
FTE2		X			X
FTE3			X		X
FTE4				X	X
FTE5	X				
FTE6		X			
FTE7			X		
FTE8				X	

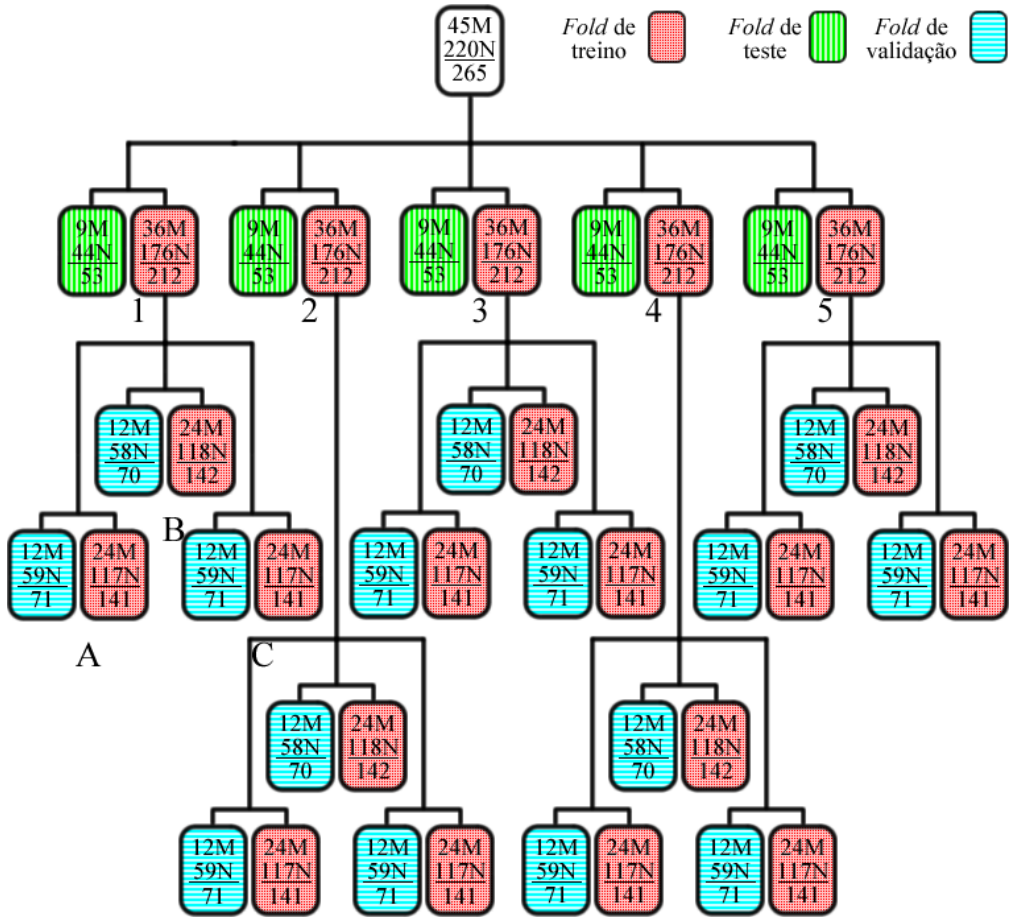
Na Etapa 3, cada variação da Tabela 2 é avaliada na diferenciação entre lesões M e N das bases da Tabela 1. Os desempenhos dessas configurações são comparados para identificar os melhores classificadores, os quais são confrontados com a literatura.

Avaliação Experimental

As medidas de avaliação empregadas são Acurácia (Ac), Sensitividade (Sen), Especificidade (Esp) e F1 *score* (F1). Essas medidas tradicionais são descritas em [2] e definidas no material suplementar (https://dx.doi.org/10.5281/zenodo.11373829).

A análise das variações de rede da Tabela 2 é apoiada pela Validação Cruzada Estratificada (VCE) [22]. Cada conjunto de imagens é dividido em cinco pares de *folds* de treino e de teste, aproximando a distribuição de classes. Cada partição de treino é dividida em três pares de *folds* de treino e de validação. Na Figura 1 esse processo é ilustrado, indicando o número de imagens com lesão M e N e o total de imagens por partição.

Figura 1 – Número de imagens com lesão maligna e não maligna em cada *fold* da base ISIC



VCE também contribuiu na estimativa do metaparâmetro quantidade de épocas para treinar cada configuração de AP a ser avaliada nos *folds* de teste, como ilustrado em [2] e nas partições da Figura 1. Para tanto, para cada variação, uma rede sobre o *fold* de treino A é inicialmente gerada. A acurácia dessa RNP em 100 épocas é examinada na partição de validação correspondente. O mesmo procedimento é realizado para B e C. Após, o número de épocas que levou à mais alta Ac média dentre A, B e C é adotado para inferir uma rede sobre o *fold* de treino 1, a qual é aplicada na partição de teste correspondente. Esse mesmo processo é realizado nos pares de *folds* 2, 3, 4 e 5.

Os *folds* da VCE são processados três vezes para cada variação de AP, e seus resultados são submetidos ao Teste de Normalidade (TN) de Shapiro-Wilk no Prism. Dependendo do resultado desse teste e do número de configurações (colunas) testadas, um teste paramétrico ou não paramétrico específico é conduzido no software.

A taxa de aprendizado é definida como $1x10^{-5}$ para FTE1, FTE2, FTE3 e FTE4. Nas demais variações são adotadas taxas de $2x10^{-5}$ e de $1x10^{-5}$, respectivamente, no quarto e no sexto procedimentos do processo de ajuste fino. A adaptação dos valores de TA com gradiente descendente é efetuada pelo otimizador *Root Mean Squared Propagation* (RMSProp) [8]. O tamanho do lote para teste iguala o número de imagens no *fold* de teste, e os tamanhos para treino e validação são fixados em uma imagem.

Representações visuais aprendidas por variações de interesse são obtidas pelo Grad-CAM [8,23], o qual realça regiões da entrada que uma rede associa à classe predita.

Resultados e Discussão

A seguir, são apresentados e discutidos os resultados sob duas perspectivas.

Comparação entre Configurações Desenvolvidas neste Trabalho

Nas Tabelas 3 e 4 são exibidos os valores médios (e desvios padrão) de acurácia, sensitividade, especificidade e F1 obtidos pelos modelos gerados nos conjuntos D104 e PH2. Os resultados alcançados na base ISIC estão no material

suplementar. As maiores médias estão em negrito. Todas as Ac médias superaram o CECM correspondente.

Tabela 3 – Configurações avaliadas neste trabalho no conjunto de dados D104

	FTE1	FTE2	FTE3	FTE4	FTE5	FTE6	FTE7	FTE8
Ac	**0,8787** (0,1471)	0,8600 (0,1499)	0,7781 (0,1527)	0,6271 (0,1121)	0,8730 (0,1742)	0,7824 (0,1312)	0,7283 (0,1345)	0,5605 (0,0899)
Sen	**0,9389** (0,1156)	0,9044 (0,1938)	0,8989 (0,1143)	0,7200 (0,1623)	0,9222 (0,1390)	0,8989 (0,1186)	0,8189 (0,1984)	0,6400 (0,2123)
Esp	0,8030 (0,2344)	0,8007 (0,2459)	0,6348 (0,3135)	0,5200 (0,2533)	**0,8074** (0,2983)	0,6370 (0,3212)	0,6141 (0,1919)	0,4667 (0,2461)
F1	**0,9000** (0,1217)	0,8736 (0,1526)	0,8251 (0,1113)	0,6778 (0,0977)	0,8969 (0,1422)	0,8277 (0,0907)	0,7631 (0,1355)	0,6035 (0,1235)

Tabela 4 – Configurações avaliadas neste trabalho no conjunto de dados PH2

	FTE1	FTE2	FTE3	FTE4	FTE5	FTE6	FTE7	FTE8
Ac	0,9700 (0,0592)	0,9750 (0,0401)	0,9450 (0,0493)	0,8900 (0,0549)	0,9817 (0,0347)	**0,9833** (0,0349)	0,9467 (0,0581)	0,8567 (0,0504)
Sen	0,9000 (0,2596)	0,9250 (0,1035)	0,7750 (0,2372)	0,4917 (0,2733)	0,9333 (0,1238)	**0,9500** (0,1035)	0,7833 (0,2521)	0,3250 (0,2664)
Esp	0,9875 (0,0284)	0,9875 (0,0284)	0,9875 (0,0259)	0,9896 (0,0152)	**0,9938** (0,0129)	0,9917 (0,0186)	0,9875 (0,0284)	0,9896 (0,0193)
F1	0,8935 (0,2615)	0,9370 (0,0980)	0,8298 (0,1720)	0,5873 (0,3182)	0,9498 (0,0957)	**0,9567** (0,0904)	0,8312 (0,2136)	0,4110 (0,3288)

A configuração FTE6 levou à maior acurácia média dentre todos os conjuntos, além de ter alcançado a maior sensitividade e F1 médias no PH2 – Tabela 4. Isso sugere que empregar uma TA maior ($2x10^{-5}$) para treinar a rede personalizada, e uma taxa menor ($1x10^{-5}$) para efetuar o treino complementar, pode levar a desempenho competitivo em bases pequenas. Em particular, a alteração realizada no parâmetro possibilitou acelerar a convergência e modificar mais as representações de imagens aprendidas no primeiro processo de treinamento, como observado anteriormente em dermoscopia [2] e outros domínios [8]. Ainda no PH2, quando se comparam apenas configurações que descongelam a mesma quantidade de blocos, se nota que outros modelos, com definição de TA idêntica à de FTE6, atingiram Ac média melhor do que a obtida pelas variações de rede com valor único de TA: (1) FTE5 superou FTE1 e (2) FTE7 ultrapassou FTE3.

FTE5, que também possui troca na taxa de aprendizado, levou (1) às melhores Sen e F1 médias no conjunto ISIC – Tabela 1 do suplementar –, no qual predominam exemplos N, e (2) à melhor Esp média no D104 – Tabela 3 –, em que imagens com lesão M formam maioria. Portanto, usar TA distintas pode contribuir no desempenho, inclusive em medidas que valorizam a predição da classe minoritária das bases citadas.

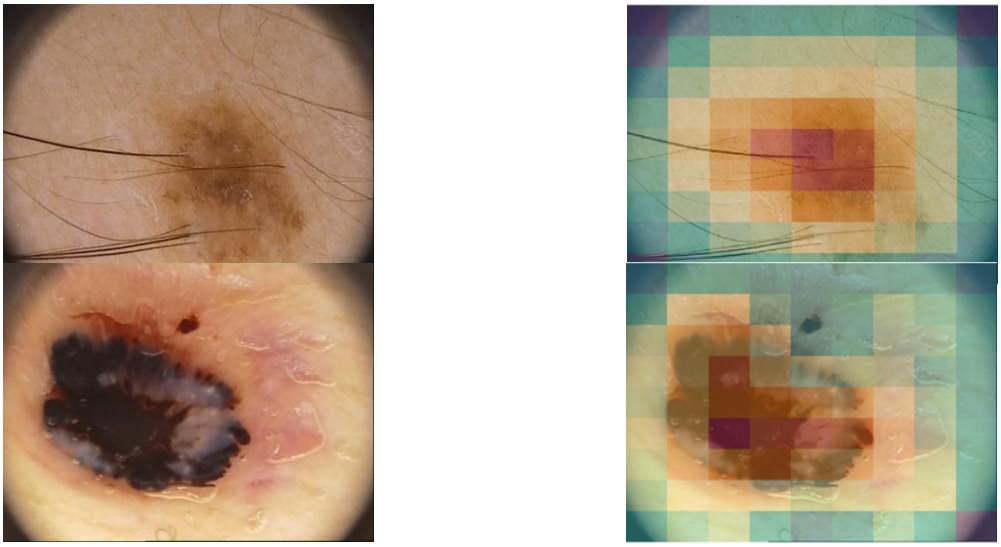
Por sua vez, FTE1 contribuiu na obtenção da melhor Ac média nas bases ISIC e D104 ao adotar uma única TA: $1x10^{-5}$. Ainda no D104, essa configuração foi associada às melhores Sen e F1 médias. Um aspecto comum entre FTE1 e FTE5 é que ambas descongelam camadas a partir do Bloco *B*=6. Logo, elas foram as configurações menos conservadoras deste trabalho, oferecendo mais camadas para especialização dos pesos correspondentes para dermoscopias. Ao mesmo tempo, ambas mantêm congeladas as camadas dos cinco blocos mais inferiores, as quais aprenderam atributos genéricos do grande repositório ImageNet que se mostraram úteis na classificação em pequenas bases dermoscópicas. Convém acrescentar que outros membros da família EfficientNet, com variações distintas, mais ou menos conservadoras que FTE1 e FTE5, foram avaliados em conjuntos dermoscópicos maiores que os deste trabalho [11-13].

Para complementar a análise, aplicou-se o teste não paramétrico de Kruskal-Wallis (nível de significância α=0,05), dado que valores das quatro medidas de avaliação não passaram pelo TN. Exceto pelos casos inerentes à Esp em PH2 e ISIC, a hipótese nula de que o desempenho dos preditores é equivalente foi rejeitada nos testes realizados. Conforme pós-teste de Dunn, notou-se que FTE1, FTE2 e FTE5 foram significativamente superiores a FTE8 nas quatro medidas no D104. Além disso, essas três redes e FTE6 foram estatisticamente superiores a FTE8 em todos os testes com rejeição da hipótese nula para PH2 e ISIC. FTE7 também foi significativamente melhor que FTE8 em Ac, Sen e F1 no PH2. Ainda nessas medidas e nesse conjunto, FTE1, FTE2, FTE5 e FTE6 foram estatisticamente melhores que a configuração FTE4.

Na Figura 2 é exibida uma amostra com duas lesões do PH2 e as representações Grad-CAM correspondentes obtidas após aplicar FTE6 nas imagens e

obter predições corretas. FTE6 foi escolhida por ter levado à melhor Ac média encontrada neste trabalho.

Figura 2 – Dermoscopias do PH2 e representações Grad-CAM correspondentes para FTE6. A primeira e a segunda linhas contém, respectivamente, um caso maligno e um não maligno



Ao sobrepor os mapas de calor Grad-CAM sobre as dermoscopias, observa-se que a RNP conseguiu identificar regiões das lesões e de suas vizinhanças com alguma importância na diferenciação das classes. Embora a explicabilidade das decisões tomadas por redes neurais ainda precise ser mais transparente a profissionais da saúde, a técnica usada é promissora [5]. Uma alternativa que pode ser investigada no futuro é a combinação de múltiplos modos de explicação das decisões, como recomendado em [23].

Comparação de Configurações Desenvolvidas em Relação à Literatura

Na Tabela 5 são exibidos os melhores resultados identificados nas menores bases deste trabalho – D104 e PH2 –, e em referências recentes. As maiores médias estão em negrito. As variações FT1 e FT2 são de um trabalho anterior [2] dos autores deste estudo. Métodos de outros pesquisadores investigados em mais de uma base

são representados na tabela pelos resultados do conjunto de dados em que o melhor desempenho foi obtido.

Tabela 5 – Resultados das melhores configurações deste trabalho e de métodos relacionados. Na última coluna é indicado o número de imagens da base de dados associada aos resultados

	Método	Ac	Sen	Esp	F1	# imagens
Este trabalho	FTE6	0,9833	0,9500	**0,9917**	0,9567	200
	FTE1	0,8787	0,9389	0,8030	0,9000	104
Literatura	FT1 [2]	0,9750	0,9167	0,9896	0,9303	200
	FT2 [2]	0,9202	0,9311	0,9044	0,9338	104
	[1]	0,9800	–	–	–	1.006
	[5]	**0,9923**	0,8900	0,9900	0,7500	58.032
	[7]	0,9870	0,9800	0,9700	0,9300	3.297
	[3]	0,9883	**0,9883**	0,9883	**0,9883**	10.015
	[6]	0,9800	0,9810	0,9750	0,9870	200
	[9]	0,8500	0,9310	–	0,8090	9.025

Discussões sobre a tabela deveriam considerar que (1) bases, estratégias de avaliação e abordagens de aprendizado podem variar entre artigos, e (2) a média e ou desvio padrão de alguma medida não estão publicados em várias referências. Em todo caso, a comparação realizada nesta seção estima quão promissoras são FTE6 e FTE1.

FTE6 foi comparável aos melhores métodos em termos de Ac, Esp e F1, atingindo respectivamente o quarto, o primeiro e o terceiro melhor resultado nessas medidas. Isso é relevante, pois FTE6 lidou com PH2, um dos menores conjuntos. FTE6 obteve esse desfecho em três execuções em *folds* da VCE, o que é útil devido ao não-determinismo de alguns procedimentos no treino da RNP. Por sua vez, os três estudos com Ac superior realizaram um único treino por *holdout* [22] em bases com no mínimo 3.297 imagens [3,5,7].

Quando se consideram somente os desempenhos no PH2, FTE6 superou FT1 [2] em todas as medidas, e EfficientNet-B0 [6] em Ac e Esp. Convém destacar que EfficientNet-B0 empregou em [6] recursos adicionais não explorados nesse trabalho: (1) validação cruzada diferenciada, na qual o conjunto HAM10000 com 10.015 dermoscopias foi usado para treinamento, enquanto que PH2 foi adotado apenas para teste, (2) pré-processamento de imagens com identificação e remoção de pelos, e (3) integração entre características extraídas pela rede com atributos projetados por

humanos (*handcrafted*). A aplicação dessas ideias no futuro pode enriquecer as configurações do presente estudo. Além disso, embora a substituição no futuro de coleções pequenas, como PH2, por HAM10000 ou outra grande base dermoscópica tenha como potencial deficiência o aumento no custo computacional do ajuste fino e a necessidade de investigação de novas variações arquiteturais no método deste trabalho, acredita-se que o desempenho resultante possa ser superior ao já obtido, ao reduzir riscos de sobreajuste (*overfitting*).

Comparações diretas FTE6 *vs.* FT1 e FTE1 *vs*. FT2 foram possíveis porque os resultados para cada *fold* da VCE de [2] estavam disponíveis. Conforme teste de Mann-Whitney (α=0,05), não foi encontrada diferença entre FTE6 e FT1 no PH2, nem entre FTE1 e FT2 no D104. Logo, EfficientNet-B2 e VGG16 [2] exibiram qualidade similar nos mesmos dados. Porém, EfficientNet-B2 oferece vantagens específicas [8,10]: (1) melhor desempenho no treinamento sobre ImageNet (https://keras.io/api/applications), indicando potencial para superar VGG16 no aprendizado por transferência em outros conjuntos, e (2) menos parâmetros, o que pode reduzir o risco de *overfitting* em futuros experimentos.

Considerando que D104 foi o menor repositório neste trabalho, voltado ao ajuste fino de redes para bases pequenas, foram investigadas ideias para tornar EfficientNet-B2 mais competitiva nesse conjunto. Observou-se que (1) as dimensões de várias imagens do D104 são menores do que as dimensões da camada de entrada da RNP: 260x260 pixels, e (2) todos os exemplos de PH2 e ISIC possuem dimensões maiores do que 260x260. Portanto, foi experimentada, apenas no D104, a redução da entrada de FTE1 a FTE8 para 120x120 pixels, um tamanho inferior a todas as imagens da base. Após três execuções das RNP, notou-se que os desempenhos melhoraram em algumas variações – vide suplementar. FTE5 foi o melhor classificador após redimensionamento da entrada, com valores médios Ac=0,9302, Sen=0,9611, Esp=0,8889 e F1=0,9454. FTE1 também evoluiu, atingindo Ac=0,9203, Sen=0,9556, Esp=0,8756 e F1=0,9346. Exceto por Esp, ambos modelos novos superaram as duas redes da Tabela 5 para D104.

Ao aplicar Kruskal-Wallis para a comparação de 16 modelos, de FTE1 a FTE8 sem e com redimensionamento, foi identificada diferença significativa. Porém, não foi encontrada diferença entre as melhores redes no pós-teste. Além disso, o uso de camadas de entradas maiores pode levar a predições melhores na ImageNet [10]. Logo, redimensionar as variações FTE1 a FTE8 não pareceu promissor para trabalhos futuros.

Convém justificar decisões tomadas no delineamento experimental. ImageNet tem sido frequente no aprendizado por transferência em imagens médicas. Os valores de TA empregados são os mesmos destacados para VGG [2,8] e estão próximos de valores testados em soluções derivadas da EfficientNet [5], a qual exibe um compromisso competitivo entre desempenho de classificação e número de parâmetros [10]. A adaptação da TA por RMSProp também ocorreu em [2,8], enquanto que o uso de um tamanho de lote para teste proporcional ao tamanho da partição foi adotado em [2]. Fixar o tamanho de lote para treino e validação em um reduziu a demanda por memória no processamento de *folds*. A avaliação de distintos valores no índice *B* tem sido examinada em diferentes arquiteturas para diferenciação de imagens médicas [2,24], mas ainda é incipiente na EfficientNet em geral e na EfficientNet-B2 em particular. EfficientNet-B2 foi a representante da família escolhida por exibir um compromisso razoável entre acurácia no ImageNet e custo computacional [10]. A adoção apenas de características aprendidas pela rede em si automatizou a tarefa de engenharia de atributos [8]. A VCE pode reduzir vieses associados à escolha de uma amostra de teste específica pela alternativa *holdout* [22]. O uso de valores *default* nos demais parâmetros diminuiu o número de variáveis estudadas.

Conclusão

Neste trabalho foram investigadas oito configurações inéditas da rede EfficientNet-B2 para classificar três pequenas bases dermoscópicas. A avaliação experimental destacou a variação FTE6, a qual obteve desempenho comparável à literatura ao usar taxas de aprendizado distintas e descongelar camadas a partir do

Bloco *B*=7. Trabalhos futuros incluem unir características *handcrafted* com aquelas extraídas por EfficientNet.

Agradecimentos

Ao Programa de Pós-graduação em Engenharia Elétrica e Computação (PGEEC)/UNIOESTE pelo apoio financeiro. Aos professores AIG Mendes, CV Nogueira e R Fonseca-Pinto pela colaboração na aquisição das imagens das bases D104 e ISIC.

Referências

1. Malik FS, Yousaf MH, Sial HA, Viriri S. Exploring dermoscopic structures for melanoma lesions' classification. Front Big Data. 2024;7:1366312.

2. Spolaôr N, Lee HD, Mendes AI, Nogueira CV, Parmezan ARS, Takaki WSR, et al. Fine-tuning pre-trained neural networks for medical image classification in small clinical datasets. Multimed Tools Appl. 2024;83(9):27305-29.

3. Balaha HM, Hassan AES. Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. Neural Comput Appl. 2023;35(1):815-53.

4. Instituto Nacional de Câncer (BR). Estimativa 2023: incidência de câncer no Brasil [Internet]. Rio de Janeiro: Instituto Nacional de Câncer; 2023 [citado 2024 Mai 22]. Disponível em: https://www.inca.gov.br/publicacoes/livros/estimativa-2023-incidencia-de-cancer-no-brasil.

5. Venugopal V, Raj NI, Nath MK, Stephen N. A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images. Decision Analytics Journal. 2023;8:100278.

6. Bansal P, Garg R, Soni P. Detection of melanoma in dermoscopic images by integrating features extracted using handcrafted and deep learning models. Comput Ind Eng. 2022;168:108060.

7. Hasan Rafi T, Shubair RM. A scaled-2D CNN for skin cancer diagnosis. In: Hallinan J, Chetty M, Heredia GR, et al., editors. Proceedings of the 18th IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology; 2021; Melbourne, Australia. [New York]: Curran Associates; 2021. p. 1-6.

8. Chollet F, Kalinowski T, Allaire JJ. Deep learning in R. 2nd ed. Shelter Island: Manning publications; 2022.

9. Liu XJ, Li KI, Luan Hy, Wang Wh, Chen Zy. Few-shot learning for skin lesion image classification. Multimed Tools Appl. 2022;81(4):4979-90.

10. Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. In: Chaudhuri K, Salakhutdinov R, editors. Proceedings of the 36th International Conference on

Machine Learning; 2019; Long Beach, United States. [Brookline]: [Microtome Publishing]; 2019. p. 6105-14.

11. Jaisakthi SM, Mirunalini P, Aravindan C, Appavu R. Classification of skin cancer from dermoscopic images using deep neural network architectures. Multimed Tools Appl. 2023;82(10):15763-78.

12. Tajerian A, Kazemian M, Tajerian M, Akhavan Malayeri A. Design and validation of a new machine-learning-based diagnostic tool for the differentiation of dermatoscopic skin cancer images. PLoS One. 2023;18(4):1-17.

13. Papiththira S, Kokul T. Melanoma skin cancer detection using EfficientNet and channel attention module. In: Wijayakulasooriya J, editor. Proceedings of the 16th IEEE International Conference on Industrial and Information Systems; 2021; Kandy, Sri Lanka. [New York]: Curran Associates; 2021. p. 227-32.

14. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. Version 2.14 [software]. 2023 [cited 2024 May 22]. Available from: http://tensorflow.org.

15. Lee HD, Mendes AI, Spolaôr N, Oliva JT, Sabino Parmezan AR, Chung WF, et al. Dermoscopic assisted diagnosis in melanoma: Reviewing results, optimizing methodologies and quantifying empirical guidelines. Knowl Based Syst. 2018;158:9-24.

16. Machado M, Pereira J, Fonseca-Pinto R. Classification of reticular pattern and streaks in dermoscopic images based on texture analysis. J Med Imaging. 2015;2(4):044503.

17. Argenziano G, Zalaudek I. Dermoscopy: a new perspective. Dermatol Pract Concept. 2011;1(1):57-8.

18. Boer A, Nischal K. A growing online resource for learning dermatology and dermatopathology. Indian J Dermatol Venereol Leprol. 2007;73(2):138-40.

19. Mendonça TF, Ferreira PM, Marçal ARS, Barata C, Marques JS, Rocha J, et al. PH2: A public database for the analysis of dermoscopic images. In: Celebi ME, Mendonça TF, Marques JS, editors. Dermoscopy Image Analysis. Boca Ratón: CRC Press; 2016. p. 419-40.

20. Codella N, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). arXiv: 1902.03368 [Preprint]. 2019 [cited 2024 May 22]: [12 p.]. Available from: https://arxiv.org/abs/1902.03368.

21. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted residuals and linear bottlenecks. In: Brown MS, Morse B, Peleg S, editors. Proceedings of the 31st IEEE Conference on Computer Vision and Pattern Recognition; 2018; Salt Lake City, United States. [Washington]: IEEE Computer Society; 2018. p. 4510-20.

22. Witten IH, Frank E, Hall MA, Pal CJ. Data mining: Practical machine learning tools and techniques. 4th ed. Burlington: Morgan Kaufmann; 2016.

23. Velden BHM, Kuijf HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Med Image Anal. 2022;79:102470.

24. Chougrad H, Zouaki H, Alheyane O. Deep convolutional neural networks for breast cancer screening. Comput Methods Programs Biomed. 2018;157:19-30.