

Aprendizado de máquina aplicado ao diagnóstico de câncer de pele

Tony Alexandre Medeiros Silva
*Departamento de Engenharia
Elétrica
Universidade de Brasília*
Brasília, Brasil
ORCID: 0000-0001-5895-8439

Ewerton Pacheco de Souza
*Departamento de Medicina
Universidade de Brasília*
Brasília, Brasil
ORCID: 0000-0001-9993-0520

Deborah Bambil
*Departamento de Biologia
Celular
Universidade de Brasília*
Brasília, Brasil
ORCID: 0000-0001-8307-0888

Guilherme de Souza e Cassia
*DF Star Hospital
Rede D'Or*
Brasília, Brasil
ORCID: 0000-0001-9029-4679

João Luiz Azevedo de Carvalho
*Departamento de Engenharia
Elétrica
Universidade de Brasília*
Brasília, Brasil
ORCID: 0000-0002-6485-6380

Francisco Assis O. Nascimento
*Departamento de Engenharia
Elétrica
Universidade de Brasília*
Brasília, Brasil
ORCID: 0000-0002-8217-1983

Resumo — Na última década, as imagens médicas têm demonstrado relevante papel no diagnóstico precoce de câncer de pele, notadamente, com o advento das tecnologias de inteligência artificial que aprendem a reconhecer padrões entre classes de imagens com base em imagens pré-classificadas, como por exemplo, bancos de imagens médicas contendo lesões cutâneas classificadas como malignas ou benignas. Neste estudo apresentamos os resultados de identificação e classificação de lesões de pele quanto à malignidade usando uma plataforma de aprendizado de máquina em desenvolvimento em nosso laboratório — a “FAON_Machine_Learning_Platform (FAON_MLP)” — e as redes neurais convolucionais pré-treinadas ResNet 152 V2, VGG16, VGG19 e AlexNet. Para a avaliação de desempenho dos algoritmos segundo métricas objetivas, se usou um banco de dados público com imagens contendo lesões de pele classificadas como benignas (1800 imagens) ou malignas (1497 imagens). São apresentados os resultados de desempenho dos diversos algoritmos. É discutida também uma avaliação comparativa entre as técnicas utilizadas e com outras técnicas reportadas na literatura. Os resultados se mostraram promissores na diferenciação entre lesões benignas e malignas, com grande potencial para auxiliar profissionais da área da saúde no diagnóstico precoce de câncer de pele, evitando os desdobramentos da sua não ocorrência, como procedimentos cirúrgicos, metástases e até mesmo óbito.

Palavras-chave—*Câncer de pele, aprendizado de máquina, redes neurais convolucionais, imagens dermatoscópicas.*

I. INTRODUÇÃO

Muitos grupos de pesquisa têm trabalhado para desenvolver ferramentas de aprendizado de máquina para diagnosticar lesões de pele, principalmente as malignas, como melanomas. Citem-se alguns dos trabalhos publicados sobre o assunto: Minagawa et al [1]; Dick et al [2]; e Tschandl et al [3].

O aprendizado de máquina fornece técnicas para que um modelo computacional possa ser criado a partir de dados

(“input”) com uma resposta (“output”) conhecida, ou dados de treinamento [4]. Em relação às lesões de pele, os dados de treinamento são fotografias de lesões com suas respectivas classificações.

Neste ponto, os trabalhos já realizados demonstraram que aprendizado de máquina — e sobretudo o “deep learning” — tem alto potencial de uso para diagnóstico de lesões de pele. Nesse sentido, Esteva et al em [5] demonstraram que a performance da sua rede neural convolucional apresentou performance similar à de dermatologistas no diagnóstico por meio de fotografias de lesões cutâneas.

No entanto, os trabalhos ainda não demonstraram quais, dentre todas as ferramentas de aprendizado de máquina, são as mais adequadas para realizar a tarefa de diagnosticar lesões cutâneas, sobretudo na diferenciação entre lesões malignas e benignas.

A proposta deste trabalho é apresentar nossa ferramenta de aprendizado de máquina para avaliar seu potencial de uso e também avaliar a performance de quatro redes neurais convolucionais pré-treinadas neste contexto.

Os esforços no sentido de diagnosticar o câncer de pele se justificam pelos números associados a esse grupo de doenças. Estima-se que, a cada ano, mais de duas dezenas de milhões de novos casos de câncer de pele são registrados no mundo, tornando as estatísticas globais alarmantes. Relatórios recentes mostram que de 2008 a 2018, houve um aumento de 53% nos novos casos de melanoma diagnosticados anualmente [6]. Considera-se a possibilidade de aumento na taxa de mortalidade desta doença entre as décadas de 2020 a 2030.

De acordo com a Sociedade Americana de Câncer [6], pacientes diagnosticados em estágios avançados de melanoma apresentam uma taxa de sobrevivência inferior a 14%, contudo, se o câncer de pele for detectado em estágios iniciais, a taxa de sobrevivência pode chegar a quase 97%. Isso aponta a

importância de se desenvolverem conhecimentos e tecnologias como as que apresentamos, com o objetivo de diagnosticar precocemente lesões de pele malignas.

Os dois tipos principais de câncer de pele são os carcinomas e os melanomas. Os carcinomas são mais frequentes, menos agressivos e com alto índice de cura. Frequentemente originam-se da epiderme (camada mais superficial da pele) e clinicamente se manifestam como lesões avermelhadas, lesões que escamam ou feridas que nunca cicatrizam. Os carcinomas se dividem em basocelular (cerca de 80% dos casos) e espinocelular (cerca de 20% dos casos). Os melanomas são mais raros; porém, mais agressivos e com alto risco de metástase (espalhamento para outras áreas do corpo) quando não são tratados precocemente. Originam-se dos melanócitos — células da pele que produzem a melanina (substância que dá coloração a pele) e clinicamente manifestam-se como pintas marrons, acastanhadas ou enegrecidas. As Figuras 1 e 2 exemplificam essas características clínicas. Com relação aos melanomas, é importante acompanhar a alteração de forma e de tamanho das pintas —, pois os contornos das lesões típicas são irregulares, e a área da lesão pode aumentar rapidamente. Em ambos os casos, de carcinomas e melanomas, importante agente etiológico é a radiação ultravioleta, que penetra a pele, causando danos ao ácido desoxirribonucleico (DNA, do inglês *deoxyribonucleic acid*) e potencialmente acarretando a malignização das células danificadas [7].

Existem vários instrumentos não invasivos que podem ajudar os dermatologistas no diagnóstico do câncer de pele, como por exemplo, imagens obtidas por câmeras convencionais ou telefones celulares. Essas fotos apresentam baixa qualidade. Uma qualidade de imagem significativamente melhor é fornecida por dispositivos dermatoscópicos que se tornaram um aparelho não invasivo significativo para a identificação de carcinomas, melanomas e outras feridas pigmentadas na pele [8]. Os autores de [9] avaliaram que um dermatologista experiente geralmente segue uma série de etapas, começando com a observação a olho nu das lesões suspeitas, depois aplica a dermatoscopia (aumento microscopicamente das lesões) seguida por biópsia. Essas atividades consomem um tempo relativamente considerável e pode levar o paciente a avançar para estágios posteriores da doença, dada a sua agressividade. Além disso, o diagnóstico preciso é subjetivo dependendo da habilidade clínica. Dermatologistas mais experientes (acima de 10 anos de prática clínica) tem acurácia inferior a 80% em diagnosticar corretamente o câncer de pele.

Muitas pesquisas têm buscado abordagens diversas que permitam prever um diagnóstico do câncer de pele de forma precisa e rápida, ainda no estágio inicial da doença. Neste momento, a capacidade de obtenção de imagens, o poder do processamento computacional associado à tecnologia da informação e das comunicações (TIC) abrem uma ampla gama de propostas para desenvolvimento de novas tecnologias e/ou agregação de valor tecnológico à infraestrutura disponibilizada atualmente para os cidadãos. O aprendizado de máquina é uma área do conhecimento, que estuda e tem como objetivo o reconhecimento de padrões, utilizando como base sua própria experiência. Os conhecimentos apreendidos, podem se relacionar à capacidade de resolver um problema, tomada de decisões, reconhecimento de padrões, dentre outros. Este campo

da inteligência artificial que reúne funções estatísticas e aprendizado de máquina a imagens é denominado visão computacional. Outros modelos, como o aprendizado profundo (“*deep learning*”, do inglês), têm se mostrado eficientes em diversas aplicações de classificação (diagnóstico) a partir das imagens e dos dados clínicos. Os modelos de aprendizado profundo podem resolver problemas críticos (de difícil identificação) de forma automática, são adaptáveis à uma variância temporal do problema considerado [11].



Fig. 1 Exemplo de imagens dermatoscópicas que representam carcinomas [10].



Fig. 2 Exemplo de imagens dermatoscópicas que representam melanomas [10].

Em nosso estudo, apresentamos um método promissor de baixa complexidade e não invasivo, para classificar lesões de pele a partir de imagens usando redes neurais convolucionais (CNN) nas arquiteturas ResNet 152 V2, VGG16, VGG19 e AlexNet e o algoritmo *Histogram-based Gradient Boosting Classification Tree* (HGBC). Nossos resultados mostram que a precisão do diagnóstico é significativa em uma avaliação abrangente e específica da anormalidade. A seguir será apresentada a metodologia utilizada para o desenvolvimento do trabalho, os resultados obtidos e uma discussão em relação aos mesmos.

II. MATERIAIS E MÉTODOS

Neste estudo não foi necessária a aprovação por comitê de ética, porque foi usado tão somente um conjunto de imagens livremente disponível para fins científicos. A seguir são apresentadas as etapas do desenvolvimento do trabalho.

A. Características do dataset

Apesar do *dataset* International Skin Imaging Collaboration (ISIC) Data Archive, disponível em <https://isic-archive.com>, disponibilizar imagens para competições desde 2012, utilizamos o conjunto de imagens liberadas para a competição de 2016. Os quais demonstram um dataset balanceado com um total de 3297 imagens, sendo distribuídas entre 1800 imagens benignas (carcinomas) e 1497 imagens malignas (melanomas) com dimensão de 224 x 224 pixels [10].

B. Extração de características

As imagens obtidas do dataset ISIC foram processadas em 2 (dois) momentos: inicialmente para a classificação com o algoritmo HGBC elas foram submetidas ao extrator de características desenvolvido por um dos autores desse trabalho

e disponível na plataforma FAON_MLP. O atual extrator extrai 785 atributos que envolvem características estatísticas de primeira ordem (histograma e forma), textura, padrão binário local, momentos de HU, momentos de Zernike, características de Haralik e informações do valor mínimo, máximo, média e desvio padrão para cada canal de cor vermelho/verde/azul (RGB, do inglês *red/green/blue*), tonalidade/saturação/brilho (HSV, do inglês *hue/saturation/value*) e CIELab [12]. Após analisar e verificar que alguns dos atributos produzidos são constantes, os mesmos são descartados, restando um arquivo de valores separados por vírgula (CSV, do inglês *comma-separated values*) de entrada com 631 atributos. Posteriormente, em um segundo momento, as imagens foram submetidas diretamente como entrada para as CNNs com suas respectivas arquiteturas.

C. Modelos de aprendizado de máquina:

O aprendizado de máquina tem evoluído e alcançado resultados significantes com relação ao reconhecimento de padrões. Nesse estudo, para validação dos dados e predição dos tumores malignos (melanomas) *versus* benignos (carcinomas) foram utilizadas as CNNs nas arquiteturas pré-treinadas ResNet 152 V2, AlexNet, VGG16 e VGG19 e o algoritmo HGBC. Ambas as técnicas com abordagem supervisionada com intuito de classificação.

As redes neurais bem como as redes profundas, apresentam características de modelagem não lineares e grande capacidade de classificação, além de serem bastante utilizadas em atividades de extração de características. Um importante exemplo dessa abordagem são as redes neurais convolucionais (CNN). Os autores de [13] e [14] apresentam aplicações de CNNs para extração de características e classificação de melanomas através de imagens dermatoscópicas. As CNNs subdividem-se em contínuas etapas para realizar a extração de características e aperfeiçoar a capacidade de reconhecer padrões em imagens em diversas etapas de execução. Elas possuem camadas inerentes a atividades realizadas, a camada de convolução, a camada de sub-amostragem (conhecida do inglês como *Pooling*) e a camada completamente conectada (conhecida do inglês, *fully connected*). A Figura 3 exemplifica essa arquitetura geral.

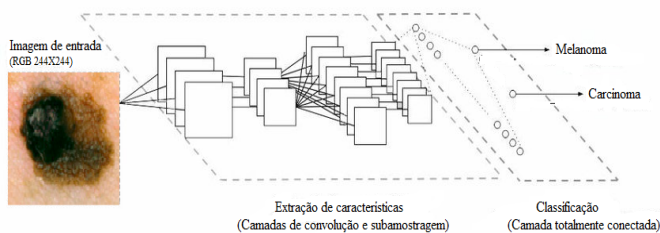


Fig. 3 Arquitetura geral de uma CNN. Adaptado de <https://www.infoq.com/br/articles/redes-neurais-convolucionais>.

O algoritmo HGBC é um modelo com suporte para valores NaN (*Not a Number*) e inspirado em LightGBM. O mesmo compõe uma classe de algoritmos de aprendizagem de *ensemble* que adicionam modelos de árvore a um *ensemble* sequencialmente. Cada modelo de árvore adicionado ao *ensemble* tenta corrigir os erros de predição cometidos pelos

modelos de árvore já presentes no *ensemble* [15]. Na prática é um processo de combinação de vários classificadores para resolver um problema complexo e melhorar o desempenho do modelo. Ele constrói um conjunto de árvores de decisão em vários subconjuntos a partir dos dados fornecidos, e ao invés de depender de uma única árvore de decisão, mescla os resultados dessas árvores com base na maioria dos votos para predizer um resultado de saída. A Figura 4 exemplifica seu funcionamento.

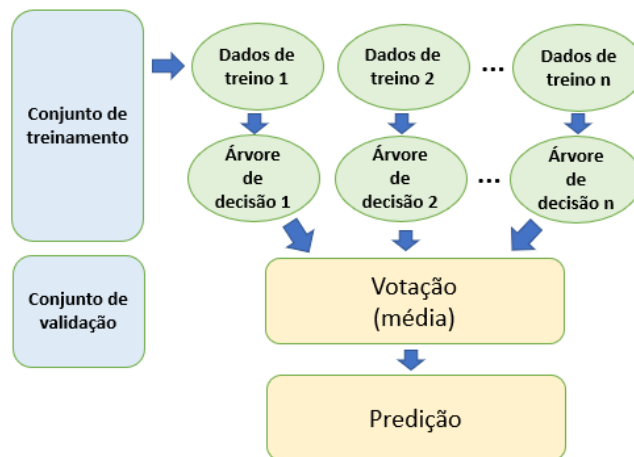


Fig. 4 Funcionamento de um algoritmo *ensemble* em árvore.

D. Classificação

O algoritmo HGBC e as CNNs foram executados por meio da plataforma FAON_MLP, versão *beta* 4. Com a utilização de computadores modernos, unidades de processamento vetorial e unidade de processamento gráfico, a plataforma aproveita muito bem os recursos de processamento paralelo e acelera os algoritmos de aprendizado de máquina. Além de implementar vários outros algoritmos de classificação, regressão e clusterização.

Conforme distribuição do *dataset* descrita no item II.a, o conjunto de dados (representado pelo arquivo CSV) de entrada para o algoritmo HGBC foi dividido em 2 subconjuntos, 80% dos dados (equivalente a 2.638 casos) para treinamento e 20% dos dados (equivalente a 659 casos) para validação.

Tanto nas etapas de treinamento quanto validação, foram aplicadas uma abordagem de validação cruzada. Essa técnica é amplamente utilizada em trabalhos nos quais o objetivo é a predição de dados. Outra técnica semelhante é o *holdout* [16], porém apresenta uma grande desvantagem, onde no momento da divisão do conjunto de dados em treinamento e validação, pode ocorrer dos dados do conjunto de treinamento serem bem parecidos, e no momento da avaliação dos dados do conjunto de validação ou teste os dados serem bem diferentes dos dados já conhecidos pelo algoritmo, e ocorrer em uma avaliação de baixa pontuação. Ainda há na literatura uma discussão sobre o melhor tamanho de divisão dos conjuntos de dados, todavia, os tamanhos 5 e 10 tem se tornado o padrão ouro em termos práticos [16]. Nesse trabalho utilizamos uma validação cruzada com *k-folds* igual a 10. Em cada execução para cada *fold* um valor de acurácia (chamado de pontuação na figura 5) é obtido no conjunto de validação. Ao final das 10 simulações, uma média das pontuações e apresentada para avaliação do

desempenho do modelo treinado. A Figura 5 apresenta como exemplo, a utilização da validação cruzada de tamanho 5.

Na implementação das CNNs, as imagens foram passadas originalmente como entrada para as redes e na primeira etapa de execução dos algoritmos foram divididas em 70% (equivalente a 2.307 casos) para treinamento, 20% (equivalente a 660 casos) para validação e 10% (equivalente a 330 casos) para teste. Executamos as CNNs por 60 épocas com uma taxa de aprendizado (*learning rate*, do inglês) igual a 0,001. Esse parâmetro é importante para realizar o ajuste da otimização do algoritmo, determinando o tamanho da etapa em cada iteração para o mínimo de uma função de perda. Foi configurado o parâmetro de parada antecipada (*early stopping*, do inglês). Inicialmente ele evita a execução do algoritmo por determinado número de épocas desnecessárias, identificando que a precisão no treinamento não está tendo melhoras, e minimiza o sobreajuste e o subajuste dos dados no treinamento.



Fig. 5 Etapas de execução da validação cruzada.

Como se trata de um aprendizado supervisionado, ao executar o treinamento dos dados pelos algoritmos HGBC e CNNs, os mesmos vão relacionar os atributos previsores (entrada) e os atributos categóricos (saída) conforme a divisão do *dataset* em treinamento, validação e teste.

III. RESULTADOS

A Tabela 1 apresenta o desempenho obtido pelos algoritmos HGBC e CNNs em suas respectivas arquiteturas na classificação dos tumores benignos (carcinomas) *versus* malignos (melanomas). Os resultados alcançados com o algoritmo HGBC e as CNNs foram no conjunto de dados de testes, porem as instâncias utilizadas não foram as mesmas, pois os dados são embaralhados na entrada do modelo. De acordo com o pré-processamento aplicado e parâmetros utilizados na etapa de classificação, ambos descritos anteriormente, avaliamos as seguintes métricas [16]: a acurácia (*Acc*), especificidade (*Esp*), sensibilidade (*Sen*), precisão (*Pre*) e f1-score (*F1*).

$$Acc = \frac{vp + vn}{vp + fp + vn + fn} \quad (1)$$

$$Esp = \frac{vn}{vn + fp} \quad (2)$$

$$Sen = \frac{vp}{vp + fn} \quad (3)$$

$$Prec = \frac{vp}{vp + fp} \quad (4)$$

$$F1 = 2 * \frac{Pre * Rec}{Pre + Rec} \quad (5)$$

$$Rec = \frac{vp}{vp + fn} \quad (6)$$

Onde, *vp* são os verdadeiros positivos, *vn* são os verdadeiros negativos, *fp* são os falsos positivos e *fn* são os falsos negativos. A métrica recall (*Rec*) foi calculada para dar origem a métrica *F1*.

A Tabela 2 apresenta a matriz de confusão do algoritmo HGBC, a Tabela 3 apresenta a matriz de confusão da CNN arquitetura VGG19, a Tabela 4 apresenta a matriz de confusão da CNN arquitetura ResNet 152 v2, a Tabela 5 apresenta a matriz de confusão da CNN arquitetura AlexNet. e a Tabela 6 apresenta a matriz de confusão da CNN arquitetura VGG16. Os valores apresentados complementam os resultados obtidos na tabela 1.

Tab. 1 Desempenho dos algoritmos HGBC e CNNs.

Algoritmo	Acc (%)	Esp (%)	Sen (%)	Prec (%)	F1 (%)
HGBC	88,56	90,72	86,24	89,51	87,83
VGG19	88,48	92,22	84,00	90,00	86,90
ResNet 152 V2	87,27	87,08	87,50	85,26	86,36
AlexNet	86,36	83,01	89,47	85,00	87,00
VGG16	76,36	90,80	60,26	85,45	70,68

Tab. 2 Matriz de confusão do resultado do algoritmo HGBC.

		Predito	
		benigno	maligno
Real	Casos de validação = 660	323	33
	benigno	42	262

Tab. 3 Matriz de confusão do resultado da CNN VGG19.

		Predito	
		benigno	maligno
Real	Casos de teste = 330	166	14
	benigno	24	126

Tab. 4 Matriz de confusão do resultado da CNN ResNet 152 v2.

		Predito	
		benigno	maligno
Real	Casos de teste = 330	155	23
	benigno	19	133

Tab. 5 Matriz de confusão do resultado da CNN AlexNet.

		Predito	
		benigno	maligno
Real	Casos de teste = 330	153	27
	benigno	18	132

Tab. 6 Matriz de confusão do resultado da CNN VGG16

		Predito	
		benigno	maligno
Real	Casos de teste = 330	158	16
	benigno	62	94

No geral, os algoritmos utilizados para a classificação dos tumores benignos *versus* malignos apresentaram resultados promissores conforme mostra a Tabela 1.

IV. DISCUSSÃO

Um grande desafio para execução de algoritmos de aprendizado de máquina é a obtenção de um *dataset* com número suficiente de casos (instâncias). O *dataset* utilizado nesse estudo forneceu imagens no formato 224x224 *pixels* que se mostraram adequadas para extração das características por meio do extrator implementado em nossa plataforma, bem como apresentou uma quantidade balanceada de casos benignos *versus* malignos, se mostrando eficaz para obtenção dos resultados obtidos conforme Tabela 1.

As respectivas métricas, acurácia, sensibilidade, especificidade, precisão e f1-score foram escolhidas por serem amplamente utilizadas em publicações de trabalhos semelhantes, permitindo uma comparação mais adequada.

Nas Tabelas 2, 3, 4, 5 e 6 é demonstrado através das matrizes de confusão, o resultado obtido através das frequências de classificação (vp, fp, vn e fn) dos algoritmos avaliando um conjunto de dados de validação/teste de novos pacientes, dos quais queremos predizer um diagnóstico assertivo.

Através dos resultados obtidos o algoritmo HGBC demonstrou capacidade de reconhecer os atributos extraídos das imagens dermatoscópicas, sobretudo pela grande quantidade de atributos extraídos pelo extrator.

As Figuras 6 e 7 mostram o comportamento das nossas CNNs com arquiteturas VGG19 e ResNet 152 V2 durante o processo de aprendizado. As curvas mostram que à medida em que a rede aprende, o erro de estimação (diagnóstico) diminui e a acurácia aumenta.

Na precisão, métrica na qual pode-se encontrar a razão entre a quantidade de casos classificados corretamente como positivos e o total dos casos identificados como positivos, os algoritmos HGBC e a CNN VGG19 apresentaram resultados superiores em comparação com os demais algoritmos implementados em nosso trabalho. Em relação à acuraria, métrica que nos diz de fato quanto dos casos foram classificados corretamente, independente da classe, os algoritmos HGBC e as CNNs VGG19 e ResNet 152 V2 apresentaram melhores resultados, porém com pequena diferença entre eles. A sensibilidade, métrica aplicada ao nosso trabalho, que representa no contexto do diagnóstico médico o percentual de acerto dos casos verdadeiros positivos foi melhor no algoritmo HGBC e na CNN ResNet 152 V2. A especificidade, outra métrica aplicada ao nosso trabalho e também muito utilizada na avaliação de trabalhos dentro de um contexto médico, e que apresenta o percentual de acerto de casos verdadeiros negativos, nos algoritmos HGBC e as CNNs VGG19 e VGG16 apresentaram os melhores resultados. Por fim, a métrica F1-Score, também aplicada ao nosso trabalho e utilizada em diversos trabalhos de aprendizado de máquina, considera a precisão e a *recall* definindo uma média harmônica entre essas duas métricas e obteve melhores resultados nos algoritmos HGBC e as CNNs VGG19 e ResNet 152 V2. Todas essas informações encontram-se na Tabela 1.

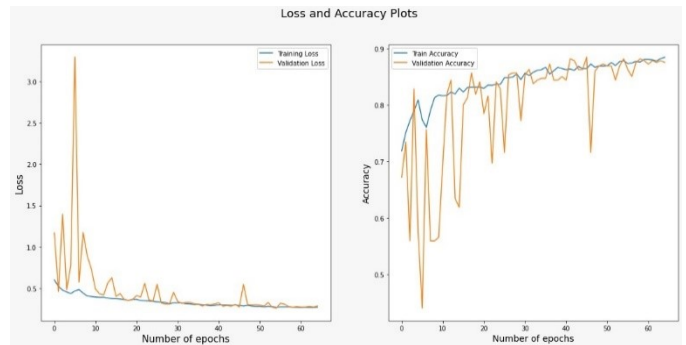


Fig. 6 Desempenho no processo de aprendizado da CNN VGG19.

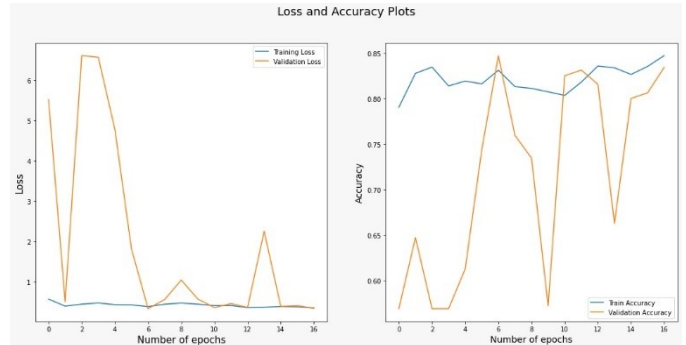


Fig. 7 Desempenho no processo de aprendizado da CNN ResNet 152 V2.

A Tabela 7 apresenta os resultados obtidos por trabalhos correlatos com utilização de CNNs com arquiteturas semelhantes às que implementamos em nosso estudo.

Tab. 7 Resultados apresentados por trabalhos correlatos no uso do aprendizado de máquina.

Referência	Algoritmo	Acc (%)
Shain et al [17]	DCNN Model	90,16
Han et al [18]	AlexNet	88,81
Guan et al [19]	VGG-16	86,09
Cargani et al [20]	DenseNet	85,25
Targ, Almeida and Lyman [21]	ResNet	85,20
Sinha and El-Sharkawy [22]	MobileNet	82,62

Pode-se notar que os resultados obtidos em nosso trabalho com os algoritmos HGBC e as CNNs com arquitetura VGG19 e ResNet 152 V2 se mostram favoráveis para utilização dos algoritmos implementados na classificação das lesões malignas da pele, superando os valores obtidos por [19], [20], [21] e [22] e ficando bem próximos dos valores alcançados por [17] e [18]. Para tal comparação foi escolhida a métrica acurácia por ser objetiva em seu resultado trazendo o número de acertos positivos pelo número total de instâncias.

A tab.7 a seguir apresenta trabalhos correlatos que também realizaram a classificação de câncer de pele, porém em *datasets* diferentes e algoritmos com abordagens próximas ao que usamos em nosso trabalho.

Tab. 8 Resultados apresentados por trabalhos correlatos na classificação de câncer de pele com mesmo dataset.

Referência	Dataset	Algoritmo	Acc (%)
Codella et al [14]	ISIC 2016	CNN - U-Net	93,4
Refianti, Mutiara and Priyandini [23]	ISIC Collection	CNN – LeNet-5	93,0
Nida et al [24]	ISIC 2016	Region based CNN (RCNN)	92,8
Gutman et al [10]	ISIC 2016	SLIC algorithm	91,6

Os resultados obtidos pelos trabalhos apresentados na Tabela 8 representam o estado da arte com respeito à utilização de aprendizado de máquina para classificação de tumores malignos *versus* benignos, principalmente quanto à detecção do melanoma, o tipo de câncer mais agressivo com conjunto de dados ISIC 2016. Os algoritmos HGBC e as CNNs com arquitetura VGG19 e ResNet 152 V2 implementados em nosso trabalho obtiveram resultados que, ao serem comparados com o estado da arte, se mostram promissores na classificação dos tumores de câncer de pele.

V. CONCLUSÃO

Foram demonstradas através deste trabalho 2 (dois) algoritmos distintos e não invasivos para classificação através de imagens, sendo que para um deles testou-se 4 (quatro) arquiteturas de redes diferentes. Ambos os algoritmos são baseados em abordagens de aprendizado de máquina para classificação de lesões malignas da pele. Todos se mostraram convenientes para realização de tal tarefa através de análise de atributos extraídos de vários descritores. Os melhores resultados foram obtidos com o algoritmo HGBC e a CNN com arquitetura VGG19. O diagnóstico adequado de câncer de pele, representa uma oportunidade para se determinar melhores decisões clínicas, cirúrgicas e terapêuticas, sobretudo se realizada nos estágios iniciais proporcionando melhor resposta ao tratamento, influenciando no tempo de sobrevida do paciente.

REFERÊNCIAS

- [1] Minagawa A. et al. Dermoscopic diagnostic performance of Japanese dermatologists for skin tumors differs by patient origin: A deep learning convolutional neural network closes the gap. *The Journal of Dermatology*. 2020. Págs. 1-5. Doi: 10.1111/1346-8138.15640
- [2] Dick V, Sinz C, Mittlbock M, Kittler H, Tschandl P. Accuracy of ϵ computer-aided diagnosis of melanoma: a meta-analysis. *JAMA Dermatol* 2019; 155: 1291–1299.
- [3] Tschandl P, Rosendahl C, Akay BN et al. Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks. *JAMA Dermatol* 2019; 155: 58–65.
- [4] Baştanlar, Y. and Özuysal, M. (2013). Introduction to “Machine Learning”. *miRNomics: MicroRNA Biology and Computational Analysis*, pp.105–128.
- [5] Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118 (2017). <https://doi.org/10.1038/nature21056>.
- [6] American Cancer Society. *Cancer facts & figures 2018*. [S.l.]: Atlanta, American Cancer Society, 2018. Disponível em: <<https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html>>. Acessado em 19/05/2021. Citado nas páginas 15, 47 e 49.

- [7] National Cancer Institute. *Melanoma Treatment (PDQ®)*. 2018. Disponível em: <<https://www.cancer.gov/types/skin/hp/melanoma-treatment-pdq>>. Acessado em 19/05/2021. Citado nas páginas 47 e 48.
- [8] Almeida, M.A.M.; Santos, I.A.X. Classification Models for Skin Tumor Detection Using Texture Analysis in Medical Images. *J. Imaging* 2020.
- [9] Massi D, Laurino M. Machine versus man in skin cancer diagnosis. *The Lancet Oncology* 2019; 20: 891–2.
- [10] Gutman, D., Codella, N., Celebi, E., Helba, B., Marchetti, M., Mishra, N., & Halpern, A. Skin lesion analysis toward melanoma detection: A challenge. In *Proceedings of the international symposium on biomedical imaging (ISBI) 2016*, hosted by the international skin imaging collaboration (ISIC) arXiv: 1605. 01397.
- [11] MURPHY, K. P. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning. [S.l.]: MIT press, 2012.
- [12] Bambil D., Pistori H., Bao F., Weber V., Alves F. M., Gonçalves E. G., Alencar F. L. F. de, Abreu U. G. P., Arruda R., and Bortolotto I. M., Plant species identification using color learning resources, shape, texture, through machine learning and artificial neural networks. vol. 40, no. 4, pp. 480–484. [Online]. Doi : <https://doi.org/10.1007/s10669-020-09769-w>.
- [13] Yu, L.; Chen, H.; Dou, Q.; Qin, J. and Heng, P. A. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE transactions on medical imaging*, IEEE, v. 36, n. 4, p. 994-1004, 2017.
- [14] Codella, N.; Nguyen, Q. B.; Pankanti, S.; Gutman, D.; Helba, B.; Halpern, A. and Smith, J. R. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*, Munich, v. 61, 2016.
- [15] Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q. and Liu T. Y., Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, Eds., vol. 30. Curran Associates, Inc., 2017.
- [16] Berrar D. Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology*, pp. 542–545. Academic Press, Oxford 2019. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>.
- [17] Shahin, I.; Nassif, A.B.; Hamsa, S.; Nemmour, N.; Hirose, K. CASA-based speaker identification using cascaded GMM-CNN 415 classifier in noisy and emotional talking conditions, *App. Soft Comp.* 2021, 103, 107141.
- [18] Han, X.B., Zhong, Y.F., Cao, L.Q., and Zhang, L.P., Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification. *Remote Sens.* 9(8):22, 2017. Article ID 848. <https://doi.org/10.3390/rs9080848>.
- [19] Guan Q, Wang Y, Ping B, Li D, Du J, Qin Y, et al. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. *J Cancer*. 2019; 10(20): 4876– 4882. <https://doi.org/10.7150/jca.28769>.
- [20] Carcagni P., Leo M., Cuna A., Mazzeo P. L., Spagnolo P., Celeste G, et al. Classification of skin lesions by combining multilevel learnings in a DenseNet architecture, *Proc. Int. Conf. Image Anal. Process.*, pp. 335-344, 2019.
- [21] Targ S., Almeida D., and Lyman K. Resnet in resnet: Generalizing residual architectures. arXiv preprint arXiv:1603.08029, 2016.
- [22] Sinha, D.; El-Sharkawy, M. Thin MobileNet: An Enhanced MobileNet Architecture. In *Proceedings of the 2019 IEEE 10th Annual Ubiquitous Computing, Electronics, New York City, NY, USA, 10–12 October 2019; Mobile Communication Conference (UEMCON): New York, NY, USA, 2019*; pp. 280–285.
- [23] Refianti R., Mutiara A.B., and Priyandini R.P. Classification of Melanoma Skin Cancer Using Convolutional Neural Network. *IJACSA*, 2019, 10 (3), 409-417.
- [24] Nida N., Irtaza A., Javed A., Yousaf M. H. and Mahmood M. T. Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy c-means clustering. *Int. J. Med. Informat.*, vol. 124, pp. 37-48, Apr. 2019.