

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Carlos Frederico Silva da Fonseca Mendes

**Utilização de redes neurais convolucionais,
descritores calculados e informações clínicas do
paciente para diagnóstico de câncer de pele**

Vitória, ES

2022

Carlos Frederico Silva da Fonseca Mendes

**Utilização de redes neurais convolucionais, descritores
calculados e informações clínicas do paciente para
diagnóstico de câncer de pele**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Informática

Universidade Federal do Espírito Santo – UFES

Centro Tecnológico

Programa de Pós-Graduação em Informática

Orientador: Prof. Dr. -Ing. Renato A. Krohling

Vitória, ES

2022

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

M538u Mendes, Carlos Frederico Silva da Fonseca, 1984-
Utilização de redes neurais convolucionais, descritores calculados e informações clínicas do paciente para diagnóstico de câncer de pele / Carlos Frederico Silva da Fonseca Mendes. - 2022.
57 f. : il.

Orientador: Renato Antônio Krohling.
Dissertação (Mestrado em Informática) - Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Pele - Doenças - Diagnóstico. 2. Redes neurais (Computação). I. Krohling, Renato Antônio. II. Universidade Federal do Espírito Santo. Centro Tecnológico. III. Título.

CDU: 004



Utilização de redes neurais convolucionais, descritores calculados e informações clínicas do paciente para diagnóstico de câncer de pele

Carlos Frederico Silva da Fonseca Mendes

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo como requisito parcial para a obtenção do grau de Mestre em Informática.

Aprovada em 07 de outubro de 2022.

Prof. Dr. -Ing. Renato Antônio Krohling
Orientador, participação remota

Prof. Dr. Vinícius Fernandes Soares Mota
Membro Interno, participação remota

A blue ink signature of Daniel Cruz Cavalieri, written in a cursive style.

Prof. Dr. Daniel Cruz Cavalieri
Membro Externo, participação remota

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Vitória/ES, 07 de outubro de 2022



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

PROTOCOLO DE ASSINATURA



O documento acima foi assinado digitalmente com senha eletrônica através do Protocolo Web, conforme Portaria UFES nº 1.269 de 30/08/2018, por
VINICIUS FERNANDES SOARES MOTA - SIAPE 1331743
Departamento de Informática - DI/CT
Em 13/10/2022 às 12:53

Para verificar as assinaturas e visualizar o documento original acesse o link:
<https://api.lepisma.ufes.br/arquivos-assinados/582433?tipoArquivo=O>



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

PROTOCOLO DE ASSINATURA



O documento acima foi assinado digitalmente com senha eletrônica através do Protocolo Web, conforme Portaria UFES nº 1.269 de 30/08/2018, por
RENATO ANTONIO KROHLING - SIAPE 1921734
Departamento de Engenharia de Produção - DEP/CT
Em 13/10/2022 às 13:06

Para verificar as assinaturas e visualizar o documento original acesse o link:
<https://api.lepisma.ufes.br/arquivos-assinados/582440?tipoArquivo=O>

Agradecimentos

Agradeço à Deus por permitir que eu chegasse até aqui.

Agradeço à minha esposa Adiseia pelo companheirismo e apoio durante o mestrado. Agradeço também aos meus pais, Sônia e Francisco, pela educação que me forneceram.

Agradeço ao meu orientador Prof. Renato A. Krohling por ter me aceitado como aluno, pelos ensinamentos e por estar sempre disponível para conversar sobre este trabalho.

Agradeço também aos colegas do LABCIN que me ajudaram com a obtenção das imagens e questionamentos durante os seminários realizados.

Resumo

O diagnóstico de lesões de pele é uma tarefa difícil devido às variações dos aspectos visuais das lesões. Como dermatologistas fazem uso de informações visuais, dados da lesão e de informações clínicas do paciente (denominados aqui como metadados clínicos), este trabalho investiga se a combinação de características extraídas com o uso de redes neurais convolucionais (CNN), características calculadas (*handcrafted features*) e metadados clínicos é capaz de melhorar o diagnóstico automático de câncer de pele. A maioria dos trabalhos sobre diagnóstico de câncer de pele na literatura, utiliza imagens dermatoscópicas e não usa metadados do paciente. Para resolver esse problema, nós utilizamos um banco de imagens clínicas coletadas através de smartphone, com informações do paciente, chamado de PAD-UFES-20. Com a arquitetura de combinação de características proposta foi verificado que a utilização dos metadados como complemento às características extraídas por redes convolucionais e as características calculadas melhorou a classificação em termos de acurácia balanceada em 7.1% para câncer e 3.2% para melanoma quando comparada com a utilização somente de características extraídas por redes convolucionais. Adicionalmente, os experimentos mostraram que combinar apenas as características calculadas da imagem com as extraídas pela CNN não melhorou os resultados, o que indica a importância do uso dos metadados clínicos para a classificação de lesões de pele.

Palavras-chaves: Diagnóstico de câncer de pele. Imagens clínicas. Informação do paciente. Fusão de características. Aprendizado profundo.

Abstract

Skin lesions diagnostic is a challenging problem due to the variety of visual aspects of the lesions. Since dermatologists make use of visual cues, lesion data and patient data (denominated here by *clinical metadata*), we investigate if the combination of features from convolutional neural networks (CNN), handcrafted features and clinical metadata can improve the performance of automated diagnoses of skin cancer. Most works on skin lesion diagnosis in the literature use dermoscopic images without clinical metadata. In order to address this problem, we used a clinical image dataset of skin lesion with patient information collected via smartphone named PAD-UFES-20. With the proposed fusion architecture we show that the results using clinical features as a complement to the CNN and handcrafted features improve the classification in terms of balanced accuracy by 7.1% for cancer and by 3.2% for melanoma as compared with only features extracted from a CNN. In addition, our findings show that combining only handcrafted features with deep features did not improve the results indicating the importance of using clinical metadata for skin lesion classification.

Keywords: Skin cancer diagnosis. Clinical images. Patient Information. Feature fusion. Deep learning.

Lista de ilustrações

Figura 1 – Dermatoscópico de imersão em óleo. (Fonte: Wikipedia)	14
Figura 2 – Diferença entre imagens da mesma lesão obtidas sem e com o uso de um dermatoscópico	15
Figura 3 – Imagens do conjunto de dados utilizado, ilustrando exemplos de semelhança interclasse em (a) e (b) e variação intraclasse em (c) e (d).	16
Figura 4 – Diagrama de blocos de um sistema tradicional de classificação de imagens.	21
Figura 5 – Neurônio biológico. (Adaptado de commons.wikimedia.org)	22
Figura 6 – Modelo de neurônio artificial (Adaptado de Haykin (2009)).	23
Figura 7 – Perceptron multicamadas.	24
Figura 8 – Rede neural convolucional e seus principais componentes.	24
Figura 9 – Exemplo da operação de convolução entre um mapa de características de tamanho 4x4 e um filtro 3x3, salto 1 e nenhum <i>padding</i> , resultando em um mapa de características de tamanho 2x2.	26
Figura 10 – Exemplo da operação <i>MaxPooling</i> em um mapa de características de tamanho 4x4 com uma vizinhança 2x2 e salto 2 resultando em um mapa de características de dimensões 2x2.	26
Figura 11 – Diferença das funções aprendidas pelas redes sem e com unidade residual.	27
Figura 12 – Arquitetura da ResNet-50 (HE et al., 2016), ilustrando as camadas e as unidades residuais. A notação $N \times$ significa que o bloco é repetido N vezes.	28
Figura 13 – Imagens clínicas do <i>dataset</i> PAD-UFES-20 coletadas por <i>smartphone</i> , mostrando os seis tipos de lesões usadas neste trabalho.	31
Figura 14 – Exemplo de codificação das características clínicas.	32
Figura 15 – Variância cumulativa explicada das componentes do PCA	32
Figura 16 – Diagrama do sistema desenvolvido, ilustrando as entradas, a saída, e as principais etapas: CNN, extração de características, redução de dimensionalidade utilizando PCA ou UMAP, combinação das características e os fatores de combinação Φ_1 e Φ_2 utilizados na determinação do número de características utilizadas nas etapas.	33
Figura 17 – Matrizes de confusão do experimento Câncer vs. Não-Câncer, mostrando a melhora no uso dos três tipos de descritores.	46
Figura 18 – Curva ROC do experimento Câncer vs. Não-Câncer.	47
Figura 19 – Matrizes de confusão do experimento Melanoma vs. Não-Melanoma.	50
Figura 20 – Curva ROC do experimento Melanoma vs. Não-Melanoma.	51

Lista de tabelas

Tabela 1	– Distribuição das doenças no banco de imagens.	30
Tabela 2	– Distribuição das doenças de pele em cada partição.	32
Tabela 3	– As sete principais doenças de pele do banco de imagens e como elas foram agrupadas em classes em cada um dos experimentos.	41
Tabela 4	– Número de características usadas para cada valor de ϕ_1 . $f_{ResNet-50}$ é calculado de acordo com a Equação 3.28.	42
Tabela 5	– Número de características usadas para cada valor de ϕ_2 . f_{imagem} é calculado de acordo com a Equação 3.29.	43
Tabela 6	– Resultados da validação cruzada para a combinação de características da CNN com as calculadas no experimento Câncer vs. Não-Câncer e PCA para a redução de dimensionalidade.	45
Tabela 7	– Resultados da validação cruzada para a combinação de características da CNN com as calculadas no experimento Câncer vs. Não-Câncer e UMAP para a redução de dimensionalidade.	45
Tabela 8	– Resultados da validação cruzada para a combinação dos três tipos de descritores no experimento Câncer vs. Não-Câncer utilizando PCA.	45
Tabela 9	– Resultados da validação cruzada para a combinação dos três tipos de descritores no experimento Câncer vs. Não-Câncer utilizando UMAP.	45
Tabela 10	– Resultados do experimento Câncer vs. Não-Câncer nos dados de teste.	45
Tabela 11	– Resultados da validação cruzada usando a combinação das características da ResNet-50 e calculadas, no experimento Melanoma vs. Não-Melanoma utilizando PCA para a redução da dimensionalidade.	48
Tabela 12	– Resultados da validação cruzada usando a combinação das características da ResNet-50 e calculadas, no experimento Melanoma vs. Não-Melanoma utilizando UMAP para a redução da dimensionalidade.	49
Tabela 13	– Resultados da validação cruzada para a combinação dos três tipos de descritores no experimento Melanoma vs. Não-Melanoma utilizando PCA.	49
Tabela 14	– Resultados da validação cruzada para a combinação dos três tipos de descritores no experimento Melanoma vs. Não-Melanoma utilizando UMAP.	49
Tabela 15	– Resultados do experimento Melanoma vs. Não-Melanoma nos dados de teste.	49

Lista de abreviaturas e siglas

PAD	Programa de Assistência Dermatológica
UFES	Universidade Federal do Espírito Santo
OMS	Organização Mundial da Saúde
CAD	Computer-aided diagnosis
RNA	Redes neurais artificiais
CNN	Convolutional neural network
SVM	Support vector machine
GLRLM	Gray level run length matrix
GLCM	Gray level co-occurrence matrix
PCA	Principal component analysis
UMAP	Uniform manifold approximation and projection

Sumário

1	INTRODUÇÃO	14
1.1	Trabalhos Relacionados	15
1.2	Motivação	18
1.3	Objetivos	19
1.4	Organização do trabalho	19
2	REFERENCIAL TEÓRICO	21
2.1	Classificação de imagens	21
2.2	Redes neurais artificiais	22
2.3	Perceptron multicamadas	23
2.4	Redes neurais convolucionais	23
2.4.1	Convolução	25
2.4.2	Pooling	25
2.4.3	Classificação	27
2.5	Redução de dimensionalidade	27
2.5.1	Análise das componentes principais (PCA)	28
2.5.2	Projeção e aproximação de variedade uniforme (UMAP)	29
3	DADOS E MÉTODOS	30
3.1	Conjunto de dados do PAD-UFES-20	30
3.2	Preparação dos dados	31
3.3	Etapas do experimento	33
3.4	Características calculadas	34
3.4.1	Assimetria	34
3.4.2	Borda	35
3.4.3	Cor	36
3.4.4	Textura	37
3.5	Características extraídas com CNN e redução de dimensionalidade	39
3.6	Fusão das características	39
4	EXPERIMENTOS E RESULTADOS	41
4.1	Experimentos	41
4.2	Métricas	43
4.3	Resultados	44
4.3.1	Câncer vs. Não-câncer	44
4.3.2	Melanoma vs. Não-Melanoma	48

5	CONCLUSÃO	52
	REFERÊNCIAS	53

1 Introdução

De acordo com a Organização Mundial da Saúde (OMS), um terço dos diagnósticos de câncer no mundo é câncer de pele (WHO, 2020). O câncer de pele do tipo não-melanoma é o quinto mais frequente no mundo e esse número provavelmente é subestimado pois o número de casos é frequentemente não rastreado pelos registros de câncer (WCRF, 2020).

Os cânceres de pele podem ser classificados em melanoma e não-melanoma. Embora raro, o melanoma é um tipo muito agressivo e é responsável pelo maior número de mortes (LINARES; ZAKARIA; NIZRAN, 2015). Ele tem origem nas células da pele e mucosa que produzem pigmentos e rapidamente pode gerar metástases. O tipo não-melanoma consiste majoritariamente de carcinoma de células basais e do carcinoma espinocelular. Embora ambos tenham baixa letalidade, o diagnóstico precoce é muito importante para prevenir danos ao tecido afetado pela doença (HELFAND et al., 2001). O câncer de pele pode ser inicialmente diagnosticado por médicos especialistas através de inspeção visual e confirmado por análise histopatológica. A inspeção visual pode ser feita sem o auxílio de instrumentos ou usando análise dermatoscópica, também conhecida como microscopia de epiluminescência (BINDER et al., 1995), uma técnica não-invasiva que permite a visualização de microestruturas na pele, invisíveis a olho nu, feita com o auxílio de um dermatoscópio, instrumento ilustrado na Figura 1. A Figura 2 ilustra a diferença entre imagens da mesma lesão, obtida sem e com o uso do dermatoscópio, respectivamente.



Figura 1 – Dermatoscópio de imersão em óleo. (Fonte: Wikipedia)

Kittler et al. (2002) comparou a acurácia da inspeção de melanoma com e sem o uso do dermatoscópio e concluiu que o seu uso melhora significativamente a acurácia do diagnóstico, mas também concluiu que a dermatoscopia não obtém resultados melhores que a inspeção sem o auxílio de instrumentos se for feita por examinadores com pouca

¹ <<https://dermnetnz.org/topics/comparative-dermoscopy-images>>

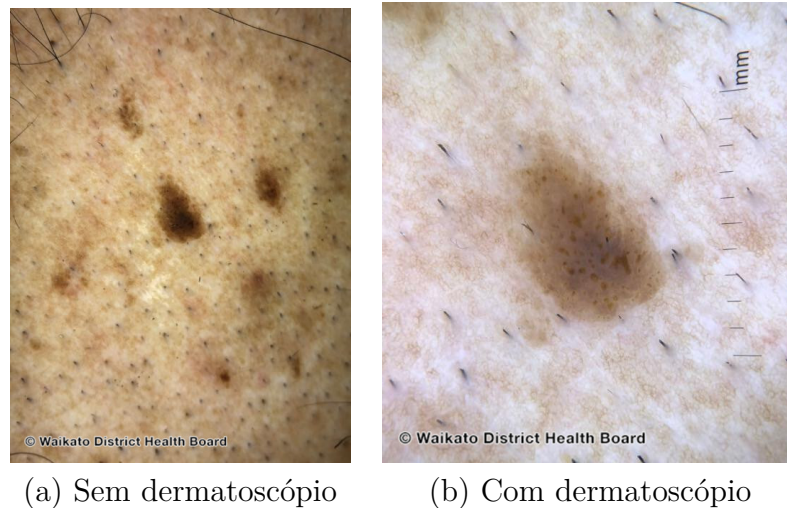


Figura 2 – Diferença entre imagens da mesma lesão obtidas sem e com o uso de um dermatoscópio (Fonte: Dermnet¹).

experiência. [Binder et al. \(1995\)](#) avaliou a análise dermatoscópica no contexto de lesões pigmentadas e concluiu que enquanto a sensibilidade do diagnóstico aumenta em 10% para especialistas, ela também diminui em 10% para não especialistas. [Ferrara et al. \(2009\)](#) mostrou que o uso de informações clínicas do paciente como idade, sexo e local da lesão pode ajudar na análise histopatológica do neoplasma melanocítico.

[Rajpara et al. \(2009\)](#) avaliou a sensibilidade e a especificidade do uso de dermatoscópio com e sem o uso de inteligência artificial para o diagnóstico do melanoma e concluiu que ambos os métodos podem ser utilizados como ferramentas de diagnóstico de melanomas.

Um dos desafios da classificação de lesões de pele é a alta semelhança interclasse e a alta variação intraclasse ([ALI et al., 2022](#)). O que significa que lesões de doenças diferentes podem ser muito parecidas, enquanto lesões da mesma doença podem se apresentar de maneiras muito distintas. As Figuras 3a e 3b mostram a semelhança entre uma amostra de melanoma e outra de nevo, enquanto as Figuras 3c e 3d mostram duas amostras diferentes de carcinoma basocelular, ilustrando, respectivamente, a semelhança interclasse e a variação intraclasse.

1.1 Trabalhos Relacionados

A maioria dos trabalhos de classificação de lesões de pele se concentram no uso de imagens dermatoscópicas pois elas possuem mais informações visuais e são frequentemente usadas por dermatologistas experientes. Ao longo dos anos, algumas regras, como a regra ABCD ([NACHBAR et al., 1994](#)) e a regra da lista de 7 pontos (*7-points checklist*) ([ARGENZIANO et al., 1998](#)) foram desenvolvidas para auxiliar os dermatologistas na interpretação das imagens e melhorar a identificação de melanoma. Pesquisas mais recentes

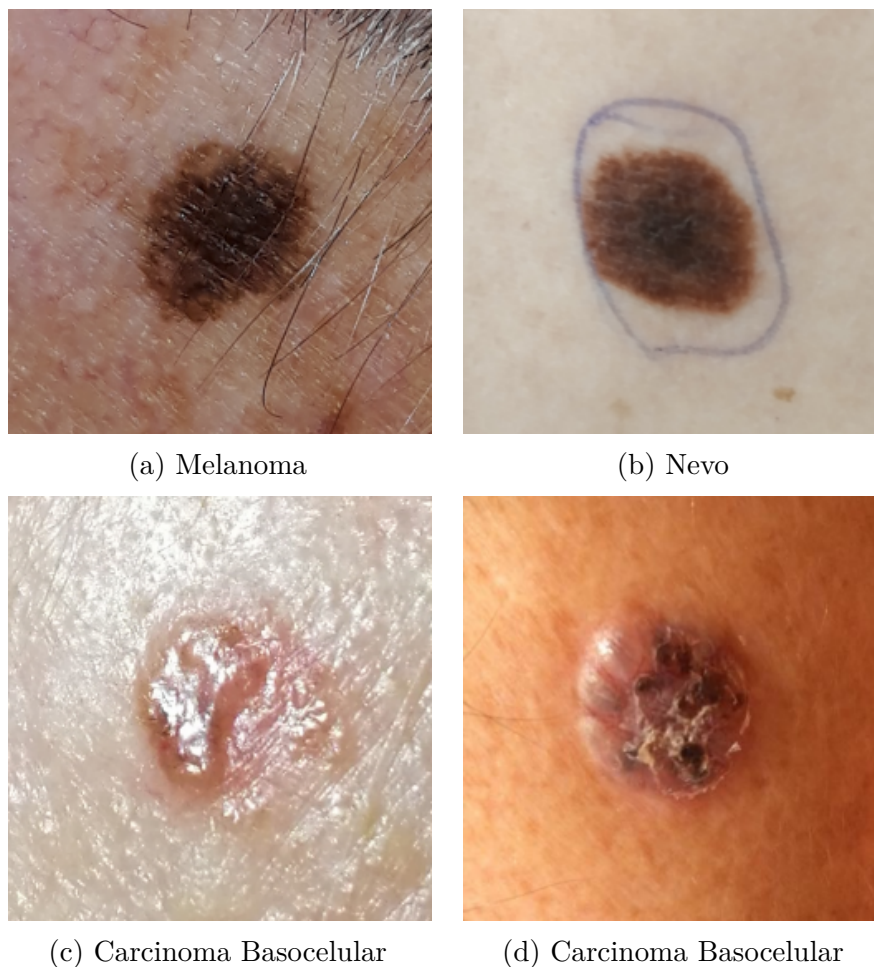


Figura 3 – Imagens do conjunto de dados utilizado, ilustrando exemplos de semelhança interclasse em (a) e (b) e variação intraclasse em (c) e (d).

em classificação automática de lesões usam métodos baseados em aprendizado profundo (*deep learning*), com poucos trabalhos explorando a combinação de características de diferentes modalidades. É importante mencionar que existe uma falta de bases de dados contendo metadados adicionais (além das imagens), o que explica a ausência desses trabalhos.

Kawahara, BenTaieb e Hamarneh (2016) usaram a rede neural AlexNet treinada no banco de imagens da ImageNet (DENG et al., 2009) para extrair características de imagens clínicas do banco de imagens de lesões de pele Dermofit. Foi demonstrado que os filtros aprendidos com a ImageNet foram suficientes para classificar as imagens com uma acurácia de 81.1%, superando um trabalho anterior no mesmo banco de imagens que obteve uma acurácia de 67.0% na classificação de 10 tipos de lesões.

Esteva et al. (2017) demonstrou a viabilidade do uso de redes neurais profundas para obter uma classificação de lesões de pele no nível de dermatologistas especialistas quando um grande número de imagens estiver disponível. Uma rede neural convolucional (CNN) GoogLeNet Inception v3 foi treinada com 129450 imagens de lesões de pele (ma-

croscópicas e dermatoscópicas). Esse é o maior banco de imagens de doenças de pele que conhecemos, porém não está disponível abertamente. A rede treinada obteve resultado comparável com o de 21 dermatologistas especialistas, em 3 diferentes modalidades de classificação: classificação de melanoma, classificação de melanoma utilizando dermatoscópio e classificação de carcinoma queratinócito.

Majtner, Yildirim-Yayilgan e Hardeberg (2016) combinaram características extraídas por redes convolucionais com dois descritores: Padrões binários locais (LBP) e RSURF. Foram utilizados dois classificadores do tipo máquina de vetor de suporte (SVM), o primeiro para classificar as características extraídas pela CNN e o segundo para classificar a concatenação dos descritores LBP e RSURF. Sendo, a saída do sistema, o rótulo atribuído pelo classificador que obteve a maior confiança. Esse método resultou em uma acurácia de 82.6% para a classificação de melanoma.

Codella et al. (2015) utilizou aprendizado profundo com codificação esparsa para aprender um dicionário de descritores para detecção de melanoma em imagens dermatoscópicas. O método obteve 93.1% de acurácia no problema de classificação melanoma vs. não-melanoma e 73.9% para melanoma vs. lesões atípicas. Foi observada uma melhora significativa quando comparada com características de mais baixo nível (*handcrafted*), sugerindo que técnicas de aprendizado não supervisionado podem ser úteis para a classificação das lesões.

Barata, Celebi e Marques (2019) fizeram uma extensa revisão de métodos de extração de características para a classificação de lesões de pele, destacando a importância de trabalhos que avaliem a sinergia de diferentes tipos de descritores.

Chatterjee et al. (2019) usou uma hierarquia de classificadores SVM e descritores do domínio espacial e de frequência para separar lesões malignas e benignas, melanocíticas e não melanocíticas. Os autores trabalharam com quatro tipos de lesões: Melanoma, Nevo, Carcinoma de células basais e Queratose seborréica. O método resultou em uma alta sensibilidade e acurácia, entretanto, eles observaram que são necessários mais testes com um conjunto maior de doenças, já que as classes escolhidas não são representativas do grande espectro de lesões possíveis.

Hameed et al. (2020) propôs uma abordagem hierárquica multinível para a classificação de lesões de pele de bancos de imagens diversos e comparou com métodos tradicionais de aprendizado de máquina baseados no uso de características de cor e textura com aprendizado profundo. O melhor resultado foi alcançado pelas técnicas de aprendizado profundo. Nenhuma combinação de características foi investigada.

Tan, Zhang e Lim (2019) apresentaram um sistema combinando GLRLM, LBP, HOG e características baseadas no método ABCD para a detecção de câncer de pele em imagens dermatoscópicas. Foi proposto seleção de características e busca de hiperparâmetros

na CNN usando o algoritmo de otimização via enxame de partículas (PSO) (KENNEDY; EBERHART, 1995) e uma coleção de classificadores para a tarefa de classificação. O sistema superou as métricas dos métodos usados como referência (*baseline*).

Em Alfed e Khelifi (2017) os autores validaram o poder discriminante dos descritores histograma de linhas (HL) e histograma de gradientes (HG) comparados as suas variações orientadas (HOG, HOL) para a detecção de melanoma. Os autores também propuseram uma abordagem baseada em um conjunto de características (em inglês: *bag of features*) para combinar cor e textura com o objetivo de melhorar a classificação de imagens dermatoscópicas e clínicas.

Yu et al. (2020) usou mapas de ativação de um conjunto de CNNs para selecionar descritores para uma posterior agregação e codificação utilizando vetores de Fisher. A estratégia aprendeu características mais robustas quando comparadas à abordagens de agrupamento. Mais recentemente, abordagens como as em Wu et al. (2020) e Zhang et al. (2019) utilizaram mecanismos baseados em atenção para melhorar a classificação e a interpretação das classificações.

Li et al. (2020) propôs um método baseado em multiplicação, semelhante a um mecanismo de atenção para combinar metadados consistindo de gênero, local da lesão e idade, com características extraídas de imagens dermatoscópicas.

Pacheco e Krohling (2020) combinaram informações extraídas de CNNs com metadados clínicos para avaliar a classificação de 6 tipos de lesões de pele utilizando um conjunto de dados de imagens não-dermatoscópicas e concluíram que a inclusão dos metadados clínicos melhorou a acurácia balanceada em aproximadamente 7%.

1.2 Motivação

Muitos países em desenvolvimento sofrem com a falta de dermatologistas e recursos médicos. Dermatoscópios não estão disponíveis em países subdesenvolvidos assim como estão em países desenvolvidos, principalmente em regiões mais afastadas das grandes cidades, contribuindo para um diagnóstico tardio do câncer de pele e, conseqüentemente, elevadas taxas de mortalidade.

Nos últimos anos, redes neurais profundas têm sido utilizadas para classificação de imagens, incluindo aplicações em análise de imagens médicas, como detecção de câncer de mama (HAMIDINEKOO et al., 2018), retinopatia diabética (ASIRI et al., 2019), entre outros. Um estudo feito por Liu et al. (2019) concluiu que o uso de aprendizado profundo pode ser no mínimo, equivalente à profissionais da saúde na detecção de doenças. Redes neurais possuem a vantagem de aprender características relevantes para a resolução do problema ao invés de precisar de um especialista no domínio para projetá-las.

Embora soluções baseadas somente no uso de CNNs mostrem resultados promissores no diagnóstico automático de câncer de pele (ESTEVA et al., 2017; BRINKER; HEKLER; ENK, 2019), elas necessitam de um grande número de imagens para serem propriamente treinadas, isso pode ser uma grande limitação no caso de imagens médicas, pois algumas doenças podem ser raras. Este trabalho tenta aproveitar o poder de generalização de redes neurais profundas, combinando-as com características calculadas (*handcrafted*) e metadados clínicos do paciente, com o objetivo de melhorar o diagnóstico em um pequeno conjunto de dados (*dataset*). Considerando que smartphones são de fácil acesso, nosso grupo de pesquisa recentemente criou um banco de imagens chamado PAD-UFES-20 (PACHECO et al., 2020), descrito na Seção 3.1 com imagens clínicas (não dermatoscópicas) e informações dos pacientes, coletado no âmbito de um projeto maior, o Programa de Assistência Dermatológica e Cirúrgica (PAD) da UFES.

1.3 Objetivos

Em um trabalho anterior realizado no laboratório (PACHECO; KROHLING, 2020), foi avaliada a importância do uso de metadados clínicos em conjunto com CNNs para a classificação de lesões de pele e concluiu-se que o uso dos metadados aumenta em, aproximadamente, 7% a acurácia balanceada em comparação com a utilização somente de CNNs. Baseado nessa premissa, este trabalho possui o seguinte objetivo:

- Adicionar uma nova fonte de informação, os descritores calculados, e investigar a combinação das três fontes de informação (KROHLING, 2020) - características extraídas com CNN, descritores calculados e metadados clínicos do paciente - no diagnóstico das lesões de pele para verificar se existe melhora na classificação no caso câncer vs. não-câncer e melanoma vs. não-melanoma.

Durante o desenvolvimento deste trabalho, foi publicado um artigo científico, referenciado como:

- MENDES, C. F. S. da F.; KROHLING, R. A. Deep and handcrafted features from clinical images combined with patient information for skin cancer diagnosis. *Chaos, Solitons & Fractals*, v. 162, p. 112445, 2022.

1.4 Organização do trabalho

O restante deste trabalho está organizado da seguinte forma:

No capítulo 2 é feita uma breve introdução ao arcabouço teórico utilizado para o desenvolvimento do trabalho. No capítulo 3 são descritos o conjunto de dados utilizados e

a metodologia. No capítulo 4 são discutidos os experimentos e os resultados. No capítulo 5 são feitas as conclusões e o direcionamento para trabalhos futuros.

2 Referencial Teórico

2.1 Classificação de imagens

A visão é o sentido mais poderoso do ser humano (HORN, 1986). A maioria das informações externas que usamos para interagir com o ambiente à nossa volta e tomar decisões, chegam, através dela, ao nosso cérebro. Sendo assim, a análise e classificação de imagens por computador é muito útil para a automatização de tarefas. A arquitetura tradicional de sistemas de classificação de imagens é mostrada na Figura 4.

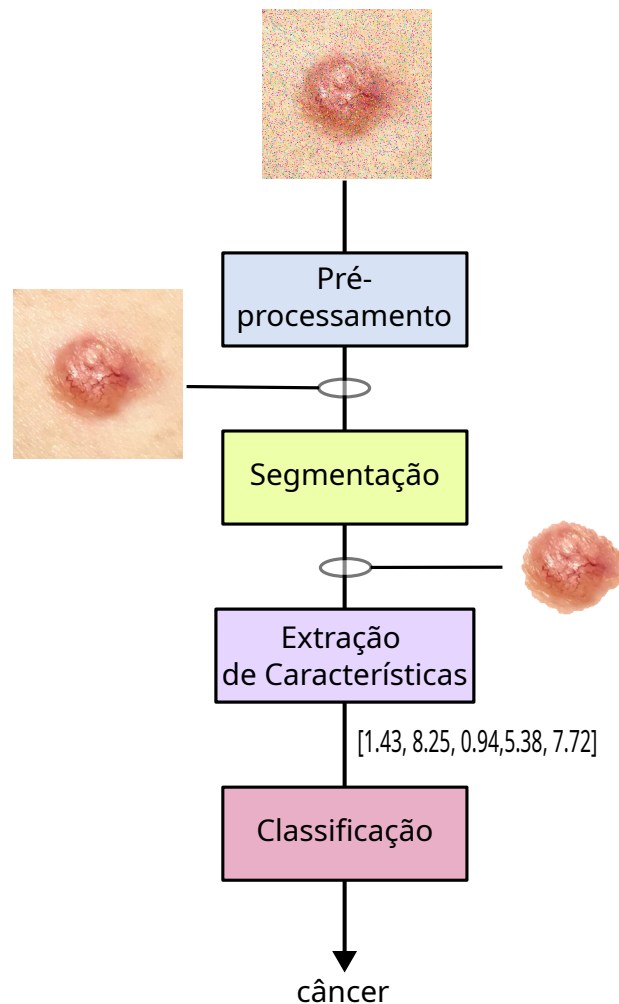


Figura 4 – Diagrama de blocos de um sistema tradicional de classificação de imagens.

O sistema da Figura 4 é dividido em 4 etapas:

Pré-processamento: Remoção de ruídos da imagem original e preparação para segmentação.

Segmentação: Identificação das regiões de interesse da imagem.

Extração de características: Obtenção de descritores para representação (quantificação) das regiões de interesse para posterior classificação.

Classificação: Classificação da imagem de acordo com os descritores obtidos na etapa anterior.

Em sistemas clássicos de processamento de imagens, a extração de características consiste de procedimentos que necessitam de conhecimento de alto nível sobre o domínio do problema para que seja feito o melhor uso da informação contida na imagem (SHEN; WU; SUK, 2017). Características (ou descritores) extraídas explicitamente utilizando o conhecimento de um especialista são chamadas de características calculadas (*handcrafted features*). Alguns descritores utilizados em doenças de pele, são, por exemplo, baseados em histogramas de cores ou informações de texturas (BARATA et al., 2014).

2.2 Redes neurais artificiais

As redes neurais artificiais são um modelo matemático de computação inspirado pelo funcionamento das redes neurais biológicas. O neurônio biológico, ilustrado na Figura 5, é a principal estrutura das redes neurais biológicas. É formado por dendritos, corpo celular e axônio. Os dendritos são os sensores responsáveis por receber os estímulos, enquanto o corpo celular reage aos estímulos recebidos propagando sua resposta para outras partes da rede pelos axônios, através de sinapses. Essa capacidade de se organizar com outros neurônios permite que o cérebro realize tarefas complexas.

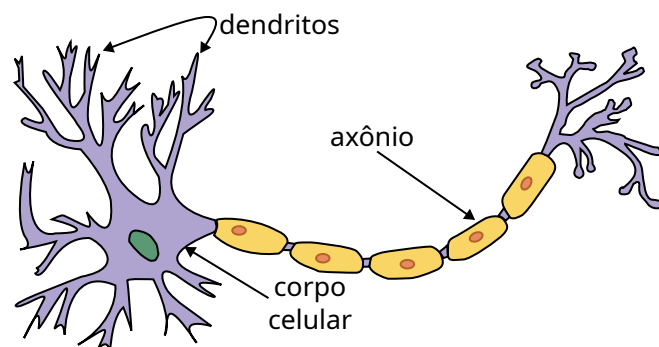


Figura 5 – Neurônio biológico. (Adaptado de commons.wikimedia.org)¹

Motivado pelo funcionamento do neurônio biológico Mcculloch e Pitts (1943) criou um modelo de computação baseado em um neurônio artificial, que foi usado por Rosenblatt (1958) na criação do Perceptron, a principal estrutura de computação das redes neurais artificiais (RNA). Rosenblatt mostrou que o perceptron era capaz de resolver problemas de detecção de padrões linearmente separáveis (HAYKIN, 2009). A Figura 6 ilustra o modelo de um neurônio artificial. Ele é composto por terminais de entrada, que assemelham-se aos dendritos, um somador, que agrega as entradas, e uma função de ativação, que determina

a resposta ao estímulo apresentado. Matematicamente, sendo as entradas do neurônio representadas pelo vetor $X = [x_1, x_2, \dots, x_n]$, $W = [w_1, w_2, \dots, w_n]$ o vetor de pesos para cada entrada, b o bias, usado para aplicar uma transformação afim na combinação linear das entradas e, a função de ativação Θ , a saída do neurônio é dada por:

$$y = \Theta(WX^T + b) \quad (2.1)$$

O vetor de pesos W pode ser interpretado como o conhecimento armazenado no neurônio. Geometricamente, são os coeficientes de um hiperplano que divide o espaço em 2 regiões diferentes, podendo assim, identificar 2 padrões diferentes.

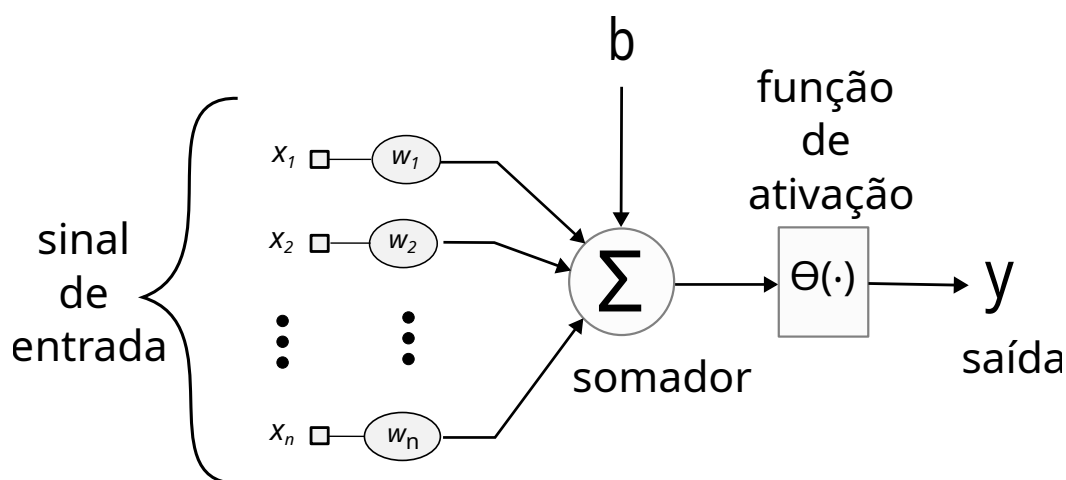


Figura 6 – Modelo de neurônio artificial (Adaptado de Haykin (2009)).

2.3 Perceptron multicamadas

Para que seja possível o reconhecimento de padrões mais complexos que os detectados pelo perceptron simples, os neurônios artificiais são organizados em camadas, onde a saída de cada neurônio da camada i , está conectada às entradas de todos os neurônios da camada $i + 1$. Essa arquitetura, ilustrada na Figura 7, é chamada de perceptron multicamadas. O vetor W de pesos de cada neurônio é obtido através de um processo de treinamento, que visa a minimização de uma função de perda. Existem vários algoritmos para ajustar os pesos de cada neurônio durante o treino, o mais utilizado é o de retropropagação (*backpropagation*) desenvolvido por Rumelhart, Hinton e Williams (1986), em que a derivada parcial da função de perda é usada para calcular a contribuição de cada peso e ajustá-lo utilizando o algoritmo de descida de gradiente (HAYKIN, 2009).

2.4 Redes neurais convolucionais

As redes neurais convolucionais (CNN) foram introduzidas por LeCun et al. (1989) como um tipo de rede neural artificial usada em visão computacional para tarefas com

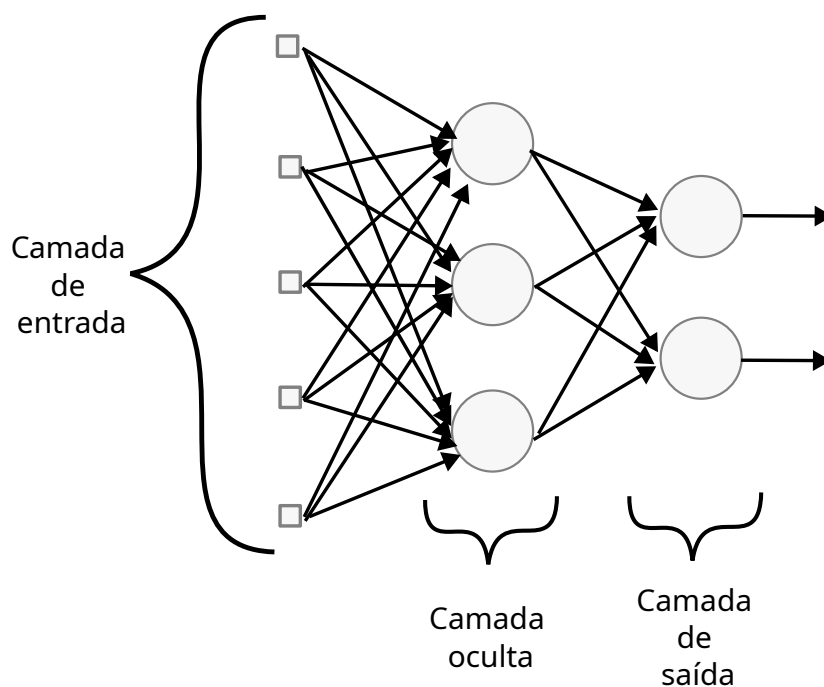


Figura 7 – Perceptron multicamadas.

imagens. Foi inspirada pelos experimentos de [Hubel e Wiesel \(1959\)](#), que mostraram que vários neurônios do córtex visual de gatos somente respondem à estímulos de uma pequena região do campo visual, enquanto a junção dos campos receptores de cada neurônio formam o campo visual completo. Embora sejam conhecidas desde a década de 1980, apenas recentemente têm sido utilizadas para a resolução em larga escala de problemas de classificação de imagens com resultados no estado da arte. Isso ocorreu, principalmente, devido aos avanços tecnológicos nas arquiteturas de processadores e memória, que possibilitaram que essas redes fossem treinadas com tempo e custos razoáveis em um grande volume de dados ([GÉRON, 2019](#)). A Figura 8 ilustra a arquitetura de uma CNN utilizada para classificação com seus principais blocos estruturais.

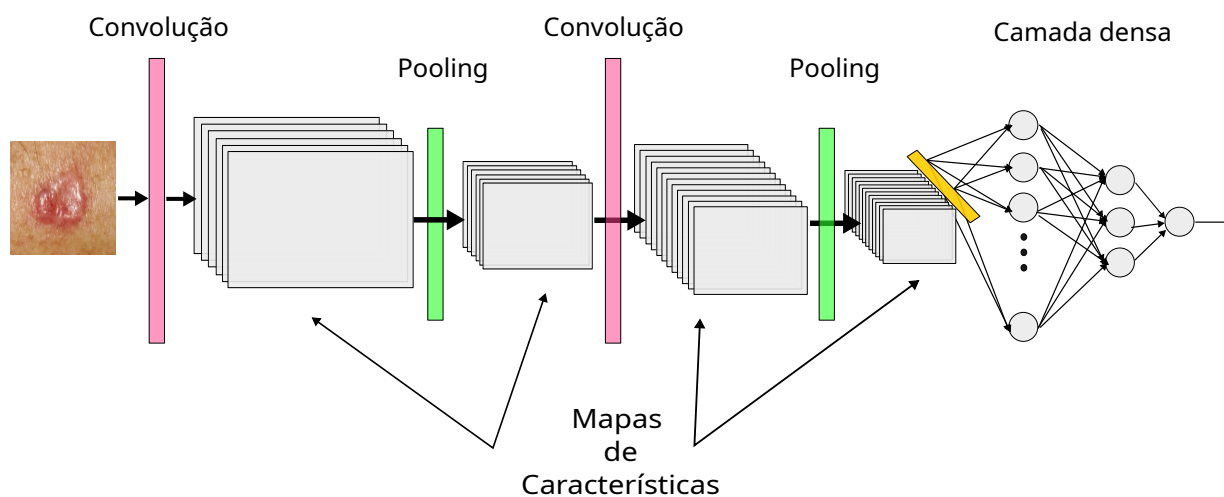


Figura 8 – Rede neural convolucional e seus principais componentes.

2.4.1 Convolução

O principal bloco de uma CNN é a camada convolucional. Cada camada convolucional pode ser vista como um conjunto de filtros que são aprendidos durante o treinamento da rede, sendo cada filtro particularmente sensível a um determinado estímulo visual. A entrada apresentada à camada convolucional passa por uma operação de convolução com cada um dos filtros e em seguida por uma função de ativação do neurônio. A saída desse processo é chamado de mapa de características (*feature map*). O conjunto de mapas de características resultantes de uma camada é usado como entrada da próxima camada. Como consequência, as características aprendidas nas camadas iniciais são mais simples que as aprendidas nas camadas finais (topo) da rede, o que permite a detecção de padrões complexos presentes na imagem de entrada. A camada de convolução possui 4 hiperparâmetros:

- **Número de filtros:** Também chamado de **profundidade**, é o número de filtros que serão usados na operação de convolução. Cada filtro dá origem a um mapa de características na saída.
- **Dimensão dos filtros:** são as dimensões das matrizes dos filtros.
- **Tamanho do passo (*stride*):** é o incremento nas direções horizontal e vertical do filtro durante a convolução.
- **Preenchimento com zeros (*zero-padding*):** É o número de linhas e colunas de zeros a serem adicionados ao redor do mapa de características antes da convolução para tornar compatíveis as dimensões do filtro e do mapa de características.

Matematicamente, o resultado da convolução é dado pela Equação 2.2 aplicada ao filtro e ao mapa de características (GOODFELLOW; BENGIO; COURVILLE, 2016):

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n), \quad (2.2)$$

onde I representa o mapa de características, K o filtro e m e n são o número de linhas e colunas do filtro, respectivamente.

Uma ilustração da operação de convolução é mostrada na Figura 9.

2.4.2 Pooling

A operação de pooling é comumente utilizada para reduzir as dimensões do mapa de características resultante das camadas convolucionais, isso possibilita uma economia de memória, processamento e a diminuição do número de parâmetros das redes. A Figura 10 ilustra a operação *MaxPooling*, que seleciona a característica com o valor mais alto em uma

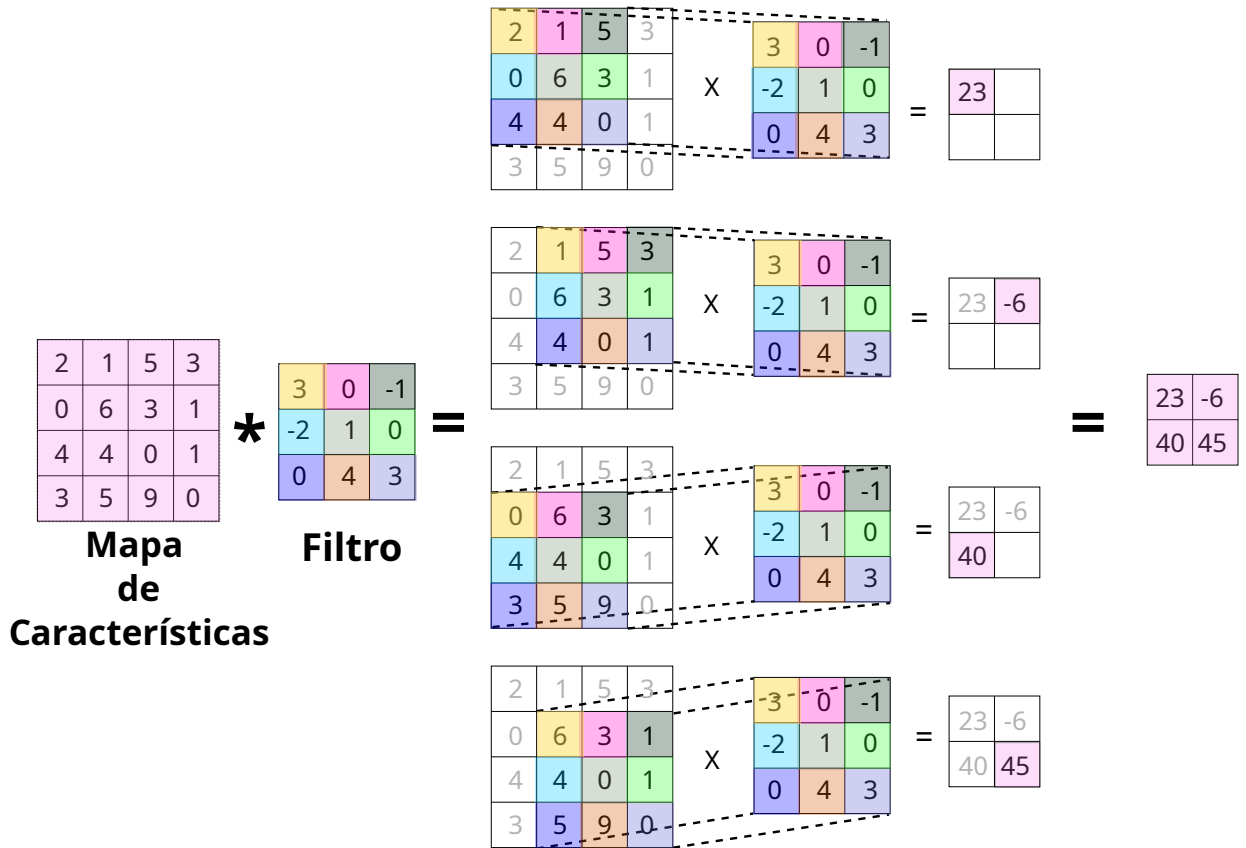


Figura 9 – Exemplo da operação de convolução entre um mapa de características de tamanho 4x4 e um filtro 3x3, salto 1 e nenhum *padding*, resultando em um mapa de características de tamanho 2x2.

vizinhança. No exemplo mostrado na Figura 10, foi utilizado uma vizinhança de tamanho 2x2 e tamanho do passo (*stride*) 2. A janela de tamanho 2x2 percorre toda a região do mapa de características em incrementos de tamanho 2, selecionando o maior valor em cada região. Além do *MaxPooling*, existem outros tipos de *pooling*, como o *AvgPooling* que calcula a média das características em uma região.

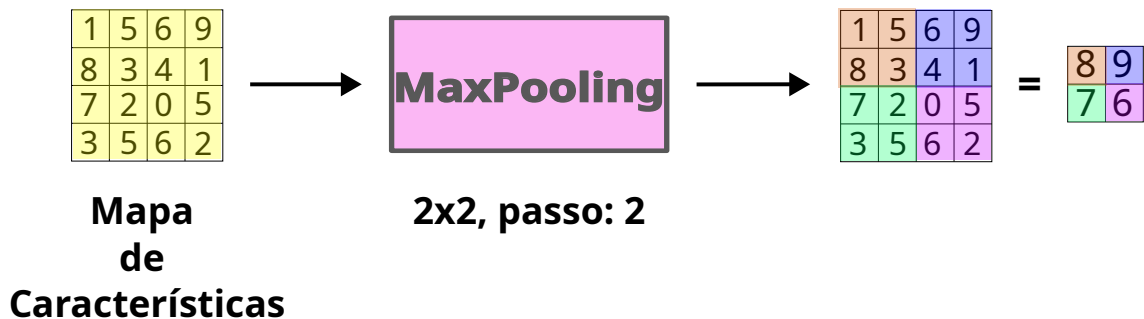


Figura 10 – Exemplo da operação *MaxPooling* em um mapa de características de tamanho 4x4 com uma vizinhança 2x2 e salto 2 resultando em um mapa de características de dimensões 2x2.

2.4.3 Classificação

Para classificar a imagem de entrada, os mapas de características da saída da última camada convolucional são usados como entrada de uma rede do tipo perceptron multicamadas (densa), para isso, os mapas de características bidimensionais são expandidos em uma dimensão e transformados em um vetor de características extraídas da CNN.

Nos últimos anos, algumas arquiteturas de redes neurais para classificação, receberam destaque na área de visão computacional, como AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), ResNet (HE et al., 2016), GoogLeNet (SZEGEDY et al., 2015), VGG (SIMONYAN; ZISSERMAN, 2015), DenseNet (HUANG et al., 2017), EfficientNet (TAN; LE, 2019) e MobileNet (HOWARD et al., 2017). Para este trabalho, a ResNet-50 foi escolhida, baseada em resultados de experimentos anteriores (PACHECO; KROHLING, 2020). A principal contribuição da arquitetura ResNet é a unidade residual. Enquanto em uma CNN sem unidade residual uma camada convolucional é treinada para aprender a função $f(x) = h(x)$, em uma rede residual, é forçada a aprender $f(x) = h(x) - x$, pois o sinal de entrada X das camadas é propagado para a saída de cada unidade residual, como ilustra a Figura 11. Isso faz com que os sinais consigam mais facilmente atravessar a rede, o que ajuda a resolver alguns problemas que podem acontecer durante o treinamento da rede neural (GÉRON, 2019). A Figura 12 ilustra a arquitetura da ResNet-50. A notação $A \times B, N$, utilizada na figura, significa uma operação de convolução com N filtros de dimensões $A \times B$.

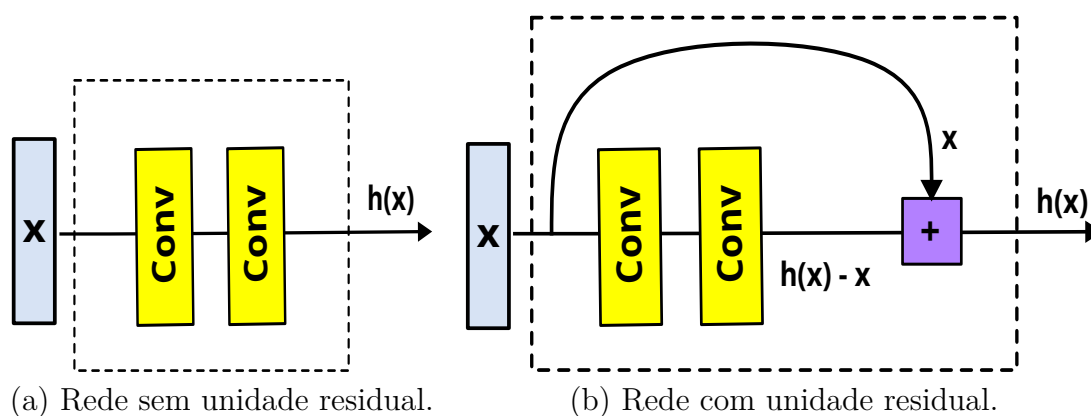


Figura 11 – Diferença das funções aprendidas pelas redes sem e com unidade residual.

2.5 Redução de dimensionalidade

Após a extração de características, é comum obter um vetor com uma quantidade de dimensões elevada, o que dificulta a classificação devido ao fenômeno conhecido como a *maldição da dimensionalidade* (BISHOP, 1995). Sendo assim, é necessário utilizar algum algoritmo para redução da dimensionalidade. Neste trabalho, foram utilizados dois algoritmos para redução da dimensionalidade, análise das componentes principais (Em

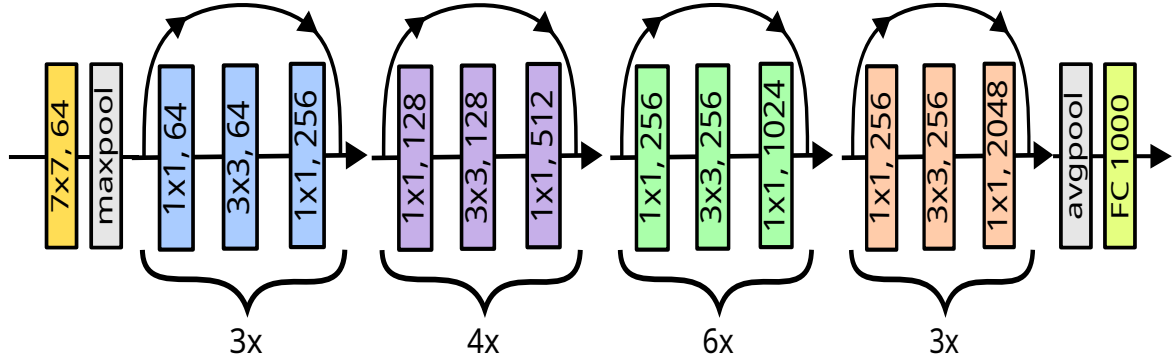


Figura 12 – Arquitetura da ResNet-50 (HE et al., 2016), ilustrando as camadas e as unidades residuais. A notação Nx significa que o bloco é repetido N vezes.

inglês: *principal component analysis* ou PCA) e projeção e aproximação de variedade uniforme (Em inglês: *uniform manifold approximation and projection* ou UMAP) brevemente descritos a seguir.

2.5.1 Análise das componentes principais (PCA)

O algoritmo mais conhecido para redução de dimensionalidade é o PCA, que consiste em achar um conjunto de vetores que indiquem as direções de maior variabilidade dos dados e então projetar os dados originais no espaço gerado por esses vetores. De acordo com Bishop (2007), sendo \mathbf{x}_n um conjunto de N observações onde \mathbf{x}_n é um vetor de tamanho D , o objetivo é projetar os vetores em um espaço de tamanho M , onde $M < D$, de maneira a maximizar a variância dos dados projetados. Simplificando o problema em uma dimensão ($M = 1$) e projetando os dados na direção do vetor unitário \mathbf{u}_1 , cada vetor \mathbf{x}_n é projetado em $\mathbf{u}_1^T \mathbf{x}_n$. A média dos dados projetados é dada por $\mathbf{u}_1^T \bar{\mathbf{x}}$, onde

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n \quad (2.3)$$

e a variância é $\mathbf{u}_1^T \mathbf{S} \mathbf{u}_1$, onde \mathbf{S} é a matriz de covariância das amostras dada por

$$\mathbf{S} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T. \quad (2.4)$$

Maximizando a variância chega-se na expressão $\mathbf{S} \mathbf{u}_1 = \lambda_1 \mathbf{u}_1$. Que mostra que a direção que maximiza a variância é dada pelo autovetor de \mathbf{S} associado ao autovalor λ_1 . O autovetor \mathbf{u}_1 é chamado de componente principal. É possível generalizar esse resultado para M componentes principais adicionando novas direções ortogonais às direções já consideradas, resultando em M autovetores $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ da matriz \mathbf{S} correspondendo aos M maiores autovalores $\lambda_1, \lambda_2, \dots, \lambda_m$.

2.5.2 Projeção e aproximação de variedade uniforme (UMAP)

O algoritmo projeção e aproximação de variedade uniforme (Em inglês: *Uniform Manifold Approximation and Projection*, abreviado por UMAP) (MCINNES et al., 2018) é uma técnica mais recente de redução de dimensionalidade baseada nas teorias de geometria Riemanniana e topologia algébrica. O UMAP é capaz de preservar estruturas globais dos dados após a projeção em espaços de menores dimensões.

Computacionalmente, o UMAP se baseia na construção de uma estrutura de grafo ponderado de k-vizinhos (em inglês: *weighted k-neighbor graph*) e de operações sobre essa estrutura. O algoritmo possui a seguinte estrutura básica:

- **Construção do grafo**

1. Construção de um grafo ponderado de k-vizinhos.
2. Aplicação de transformações nos vértices do grafo.
3. Tratamento da assimetria do grafo.

- **Organização do grafo**

1. Definição de uma função objetiva que preserve as características do grafo.
2. Encontrar uma representação em menor dimensão que otimize a função objetiva.

O algoritmo possui os seguintes hiperparâmetros:

- Número de vizinhos.
- Dimensão do espaço reduzido.
- Deparação desejada entre pontos próximos no espaço reduzido.
- Número de épocas utilizadas para otimizar a função objetiva.

Dentre as vantagens do UMAP, de acordo com McInnes et al. (2018), estão a melhor preservação da estrutura global dos dados após a redução da dimensionalidade quando comparado ao algoritmo PCA e ao t-SNE. O UMAP também é computacionalmente mais eficiente em tempo de execução quando comparado ao t-SNE. O autor não ofereceu informações sobre a comparação temporal com o PCA. Dentre as desvantagens do UMAP, estão a falta de interpretabilidade quando comparado ao PCA e também, devido às premissas assumidas pelo UMAP, ele pode achar estruturas em ruídos dos conjuntos de dados, o que pode confundir o algoritmo, principalmente quando utilizado em conjuntos de dados pequenos. O leitor interessado deve consultar McInnes et al. (2018) para mais detalhes sobre o UMAP.

3 Dados e métodos

Neste capítulo são apresentados o conjunto de dados (*dataset*) utilizado, a preparação dos dados, os descritores calculados e a descrição dos experimentos.

3.1 Conjunto de dados do PAD-UFES-20

O banco de imagens do PAD-UFES-20 (PACHECO et al., 2020) consiste de imagens clínicas obtidas com *smartphones*, diagnóstico da lesão, idade do paciente, local da lesão e a respostas (sim ou não) para as seguintes perguntas: 1) A lesão coça? 2) A lesão sangra? 3) Aumenta de tamanho? 4) Dói? 5) Muda de aparência? 6) A lesão possui elevação na pele?

Foram usados sete tipos de lesões de pele do banco de imagens: carcinoma de células basais, carcinoma espinocelular, melanoma, queratose actínica, queratose seborréica, nevo e lentigo maligno. Como o lentigo maligno pode progredir para *melanoma in situ*, foi categorizado como melanoma (KASPRZAK; XU, 2015). A Tabela 1 lista o número de imagens de cada tipo de lesão no banco de imagens utilizado. É importante observar que os dados são desbalanceados, especialmente com relação ao melanoma, já que é um câncer raro. Esse é um problema comum em aplicações de imagens médicas. A Figura 13 ilustra algumas imagens das lesões do dataset utilizado. A versão completa do dataset está disponível em <<https://data.mendeley.com/datasets/zr7vgbcyr2/1>>. O conjunto de imagens utilizadas é um subconjunto do PAD-UFES-20 para o qual haviam máscaras de segmentação feitas com o objetivo de extração das características calculadas, visto que a segmentação das lesões não está no escopo deste trabalho.

Diagnóstico	Nº de imagens
Carcinoma de células basais (BCC)	311
Carcinoma espinocelular (SCC)	101
Melanoma (MEL)	44
Queratose actínica (ACK)	330
Queratose seborréica (SEK)	135
Nevo (NEV)	145
Total	1066

Tabela 1 – Distribuição das doenças no banco de imagens.

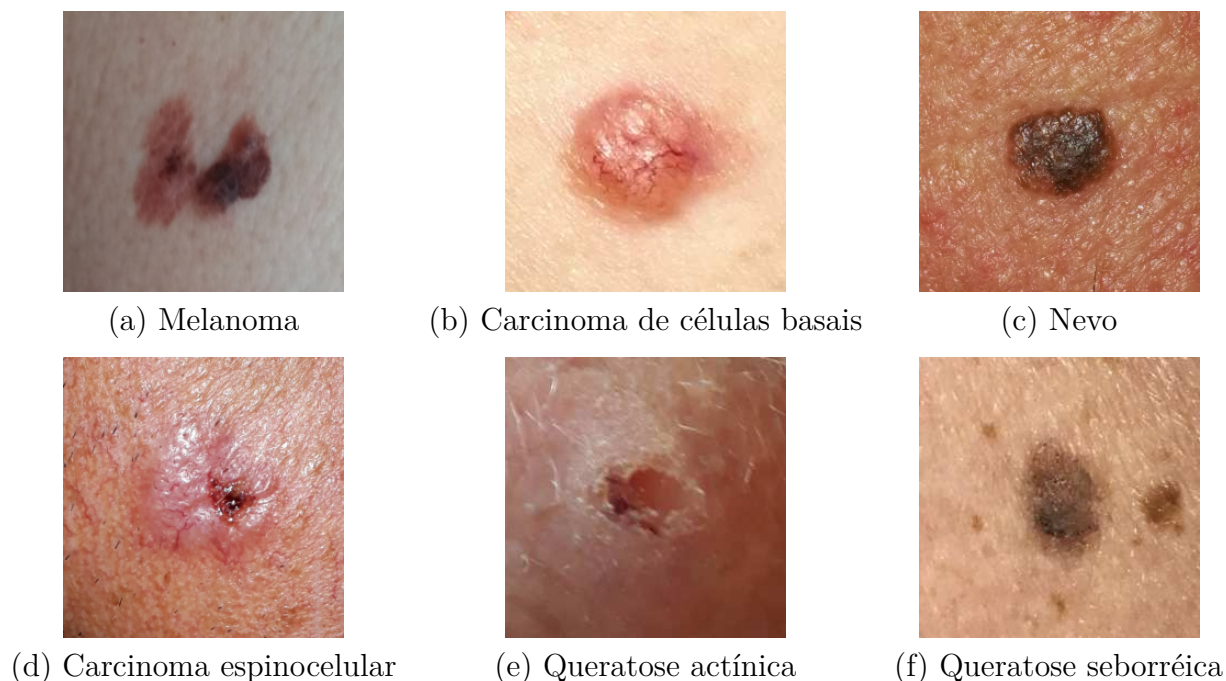


Figura 13 – Imagens clínicas do *dataset* PAD-UFES-20 coletadas por *smartphone*, mostrando os seis tipos de lesões usadas neste trabalho.

3.2 Preparação dos dados

O *dataset* foi particionado em treino, validação e teste, nas proporções de 80%, 10% e 10%, respectivamente. Como alguns pacientes possuíam mais de uma imagem no banco de dados, foram tomados cuidados para que todas as imagens de um determinado paciente apareçam em apenas uma das partições, evitando assim, viés na classificação. Também foram removidas imagens similares do mesmo paciente. As características extraídas da CNN e as calculadas foram normalizadas utilizando *z-score*, já que cada descritor possui uma faixa de valores diferentes.

As informações do paciente estão divididas em: região do corpo onde está a lesão, idade do paciente e 6 perguntas sobre os sintomas cujas respostas podem ser “sim”, codificado como 1, ou “não”, codificado como 0, resultando em 8 descritores. Para as regiões do corpo onde se encontram a lesão, foram consideradas 15 áreas principais, resultando em um vetor *one-hot* de comprimento 15. Para a idade do paciente, foi escolhido agrupar os valores em 3 grupos: jovem, adulto e idoso, resultando em um vetor *one-hot* de comprimento 3. Sendo assim, o comprimento total do vetor de informações do paciente é 24 (15 + 3 + 6). A Figura 14 ilustra um exemplo da codificação das características clínicas.

A Tabela 2 lista a distribuição das lesões em cada partição do *dataset*. Para o *baseline* comparativo (valor de referência) foi utilizado o PCA para a redução de dimensionalidade, mantendo as componentes que explicam 95% da variância, o que resultou em 22 descritores, como mostrado na Figura 15.

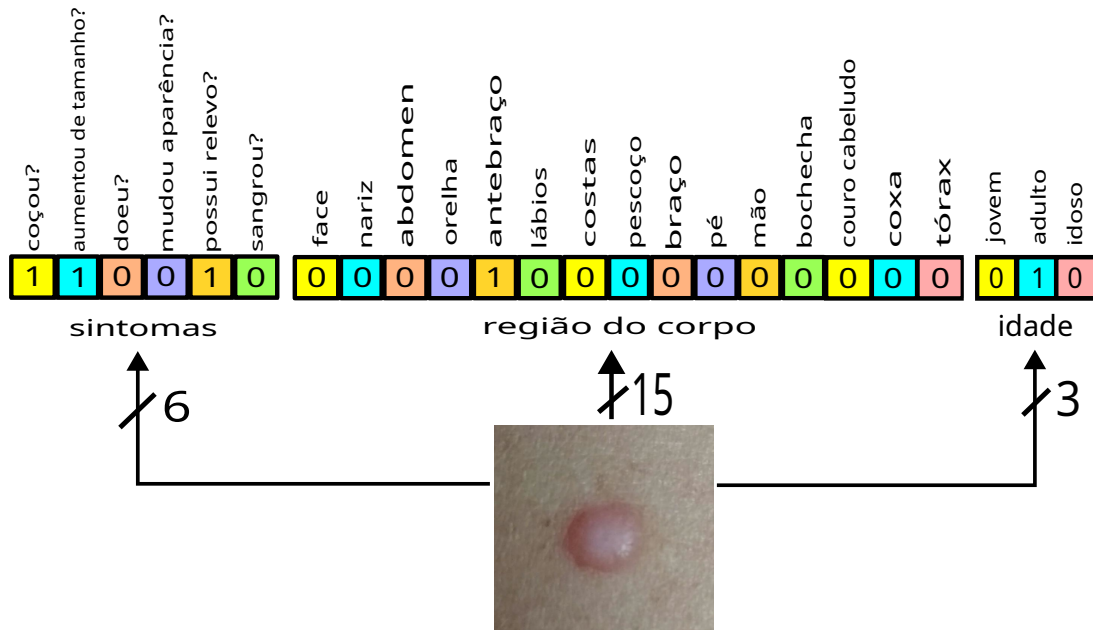


Figura 14 – Exemplo de codificação das características clínicas.

Diagnóstico	Treino	Validação	Teste	Total
Carcinoma de células basais	241	33	37	311
Carcinoma espinocelular	80	13	8	101
Melanoma	8	3	4	15
Lentigo maligno	15	7	7	29
Queratose actínica	259	37	34	330
Queratose seborréica	110	14	11	135
Nevo	111	17	17	145
Total	824	124	118	1066

Tabela 2 – Distribuição das doenças de pele em cada partição.

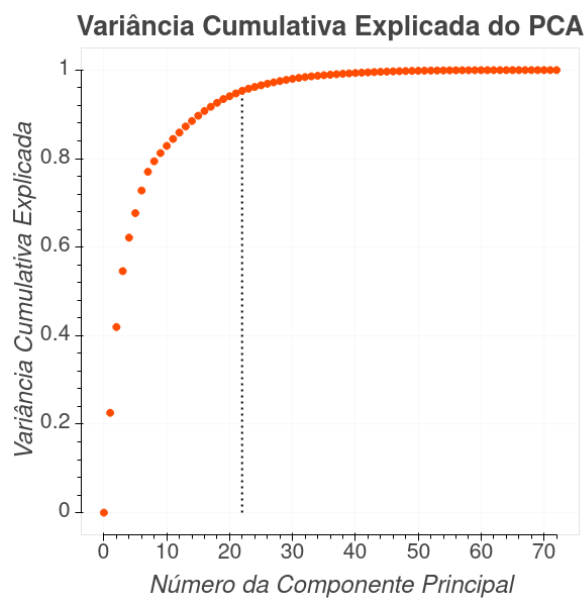


Figura 15 – Variância cumulativa explicada das componentes do PCA

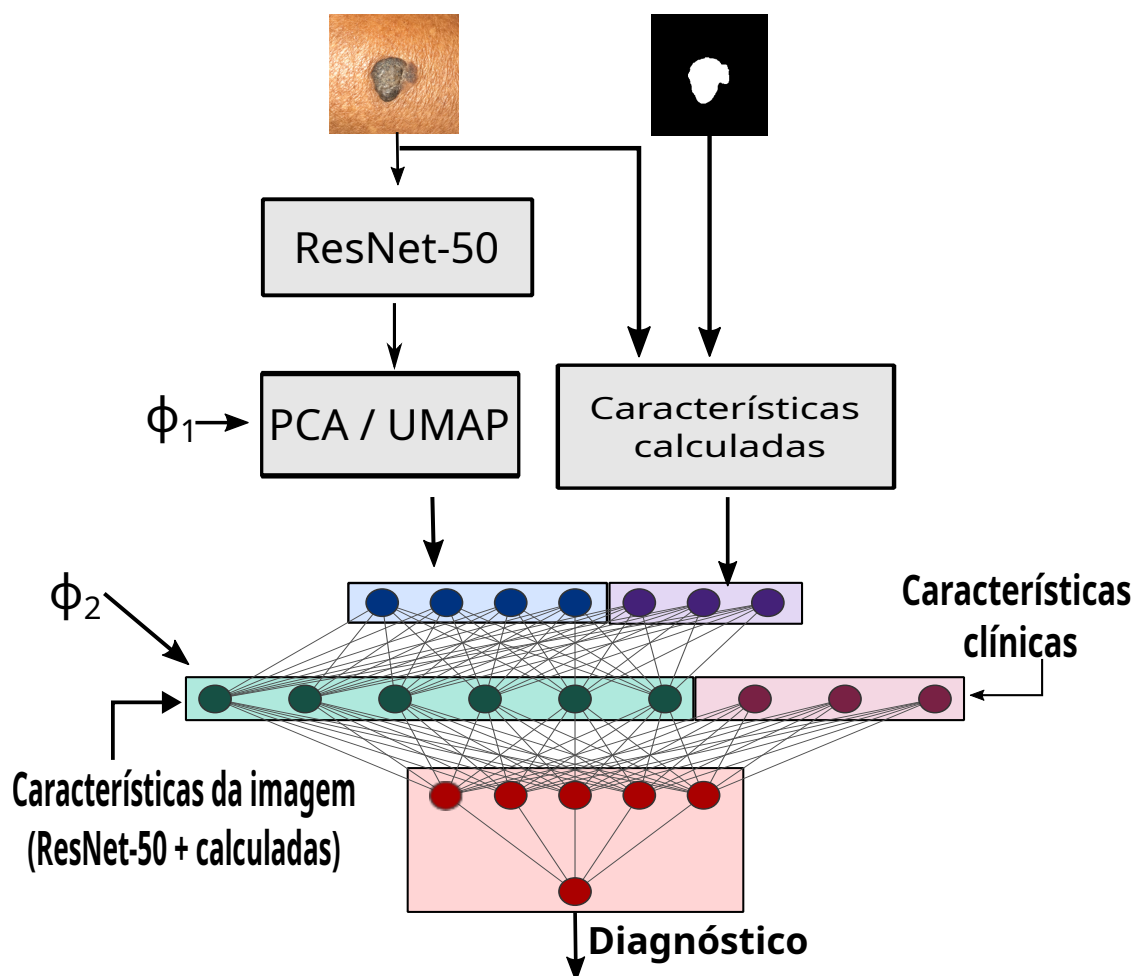


Figura 16 – Diagrama do sistema desenvolvido, ilustrando as entradas, a saída, e as principais etapas: CNN, extração de características, redução de dimensionalidade utilizando PCA ou UMAP, combinação das características e os fatores de combinação Φ_1 e Φ_2 utilizados na determinação do número de características utilizadas nas etapas.

3.3 Etapas do experimento

O método proposto pode ser dividido em quatro etapas:

1. Extração dos descritores calculados
2. Treinamento da CNN, extração das características e redução da dimensionalidade
3. Treinamento da rede neural para combinação das características
4. Classificação.

A Figura 16 mostra o diagrama de blocos do sistema desenvolvido, ilustrando as entradas do sistema, a CNN e a rede para combinação das características de diferentes modalidades, representada por retângulos coloridos. Como este trabalho se concentra na

extração de características e na classificação de imagens, a etapa de segmentação está fora do escopo, o leitor interessado pode consultar [De Angelo, Pacheco e Krohling \(2019\)](#). Assim, foi preciso utilizar máscaras feitas manualmente para segmentação de cada lesão e subsequente extração das características calculadas. As entradas do sistema são a imagem clínica da lesão, a máscara de segmentação e as informações do paciente. A arquitetura proposta não foi treinada fim a fim já que o algoritmo de redução de dimensionalidade necessita da disponibilidade prévia de todas as características. O treinamento da rede de fusão é feito após a redução da dimensionalidade do vetor de características da CNN. O processo de combinação das diferentes modalidades de descritores é uma extensão do método usado em [Pacheco e Krohling \(2020\)](#). Inicialmente, a saída da etapa de redução de dimensionalidade é concatenada com os descritores calculados, resultando nos *descritores da imagem*, que, após esta etapa, são combinados com as informações do paciente em outra camada da rede.

3.4 Características calculadas

Características calculadas são descritores baseados em um conhecimento prévio sobre o problema. Descritores clássicos usados no problema de classificação de lesões de pele são inspirados pela regra ABCD da dermoscopia para melanoma ([ARGENZIANO et al., 1998](#)), que estabelece 4 tipos de categorias de descritores: Assimetria, Borda, Cor e Dermoscópicas. A categoria de dermoscópicas se refere às estruturas que são possíveis de observar em imagens de dermatoscópio. Como neste trabalho foram usadas imagens adquiridas por *smartphone*, essas estruturas não estão presentes, sendo assim, essa categoria foi substituída por descritores de textura. Foram utilizados um conjunto de 72 descritores conhecidos da literatura com o objetivo de representar a assimetria, variação e distribuição de cor, irregularidade das bordas e textura das lesões ([ESGARIO; KROHLING, 2022](#)). A seguir cada descritor utilizado é brevemente explicado.

3.4.1 Assimetria

Os descritores de assimetria quantificam a assimetria da lesão em termos de formato e de cor.

1. Assimetria de Forma (A_{1-3})

A assimetria da forma é calculada como em [Cherifi \(2014\)](#):

$$A_1 = \frac{\alpha \cap \beta}{\alpha \cup \beta} \quad (3.1)$$

onde α e β são a máscara da lesão rotacionada de 0° e 180° , respectivamente. É a razão entre o número de pixels na intersecção das máscaras pelo número de pixels

na união das máscaras. Adicionalmente, a máscara é rotacionada para que o eixo maior da lesão fique em posição horizontal e o processo é repetido para as metades da direita e esquerda e para as metades superior e inferior da imagem:

$$A_2 = \frac{\alpha_{esquerda} \cap \alpha_{direita}}{\alpha_{esquerda} \cup \alpha_{direita}} \quad (3.2)$$

$$A_3 = \frac{\alpha_{superior} \cap \alpha_{inferior}}{\alpha_{superior} \cup \alpha_{inferior}} \quad (3.3)$$

2. **Assimetria de Cor** (A_{4-5}) A assimetria de cor é calculada como uma medida da distância entre os histogramas RGB de partes da imagem e é baseada em [Smaoui e Bessassi \(2013\)](#):

$$D(h_1, h_2) = \sum_{i=1}^n \frac{(h_1(i) - h_2(i))^2}{(h_1(i) + h_2(i))} \quad (3.4)$$

onde h_1 e h_2 são os histogramas de diferentes partes da imagem. Assim, calcula-se:

$$A_4 = D(h_{esquerda}(i), h_{direita}(i)) \quad (3.5)$$

$$A_5 = D(h_{superior}(i), h_{inferior}(i)) \quad (3.6)$$

3.4.2 Borda

Para quantificar a irregularidade da borda foram usados os seguintes descritores já conhecidos da literatura: Compacidade, Dimensão Fractal, Variância Radial, Transição de Pigmentação, Solidez e o método de Jaworek para avaliação de bordas ([JAWOREK-KORJAKOWSKA; TADEUSIEWICZ, 2014](#)).

1. **Compacidade** (B_1) Compacidade é a medida do quanto uma forma se assemelha a um círculo. Um círculo possui compacidade igual a 1. Neste trabalho, a compacidade foi calculada como em [Bhuiyan, Azad e Uddin \(2013\)](#):

$$B_1 = \frac{p^2}{4\pi\alpha} \quad (3.7)$$

onde p é o perímetro da forma e, α , a sua área.

2. **Variância Radial** (B_2)

A distância radial é a distância entre o centróide da forma e um ponto de seu contorno. A irregularidade da borda pode ser medida como a variância da distância radial ao longo da forma e é chamada de variância radial ([CHERIFI, 2014](#)):

$$B_2 = \frac{\frac{1}{N} \sum_i^N (D(p_i, G) - m)^2}{m^2} \quad (3.8)$$

onde N é o número de pixels no contorno, G é o centróide e m é a distância média da borda ao centróide.

3. **Solidez** (B_3) Solidez é a razão entre a área da lesão e a área do menor polígono convexo que envolve a região, também chamado de *convex hull*.

$$B_3 = \frac{\alpha}{\alpha_{CH}} \quad (3.9)$$

4. **Transição de Pigmentação** (B_{4-5}) A transição de pigmentação mede as transições abruptas entre a lesão e a pele saudável. Ela é calculada usando o gradiente dos pixels da borda. As características utilizadas são a média e a variância dos gradientes dos pixels da borda na imagem da pele em tons de cinza.
5. **Dimensão de Hausdorff** (B_6) A dimensão de Hausdorff é a medida da rugosidade de um objeto (forma). Foi utilizado o método de caixa de contagem para calcular a dimensão de Hausdorff do contorno da lesão, como em [Deviha \(2016\)](#).
6. **Método de Jaworek para avaliação da irregularidade de bordas** (B_7) Em [Jaworek-Korjakowska e Tadeusiewicz \(2014\)](#) foi proposto um método para medir a irregularidade de bordas em imagens dermatoscópicas de lesões. O método consiste em achar o contorno exterior da lesão, seu retângulo delimitador e a distância dos pixels da borda até o retângulo delimitador. Essa distância é suavizada para aliviar os ruídos e o número de pontos críticos dessa função é usado como descritor.

3.4.3 Cor

A variedade de cores na superfície da lesão pode diferenciar algumas doenças de pele. Foram extraídas estatísticas do histograma e uma quantidade numérica chamada de variação de cor, para descrever a distribuição de cores da lesão.

1. **Histograma** (C_{1-18})

Foram obtidos os histogramas RGB e HSV e partir deles foram calculadas a média μ , variância σ , e a distorção S (skewness), de cada canal:

$$\begin{aligned}\mu &= \sum_{i=0}^{L-1} r_i p(r_i), \\ \sigma &= \sum_{i=0}^{L-1} (r_i - \mu)^2 p(r_i) \\ S &= \sum_{i=0}^{L-1} (r_i - \mu)^3 p(r_i)\end{aligned}$$

$$C_{1-3} = (\mu(R), \sigma(R), S(R)) \quad (3.10)$$

$$C_{4-6} = (\mu(G), \sigma(G), S(G)) \quad (3.11)$$

$$C_{7-9} = (\mu(B), \sigma(B), S(B)) \quad (3.12)$$

$$C_{10-12} = (\mu(H), \sigma(H), S(H)) \quad (3.13)$$

$$C_{13-15} = (\mu(S), \sigma(S), S(S)) \quad (3.14)$$

$$C_{16-18} = (\mu(V), \sigma(V), S(V)) \quad (3.15)$$

2. Variação da Cor (C_{19-24})

A variação da cor é definida como o logaritmo da razão entre a variância σ e a média μ de cada canal das imagens RGB e HSV.

$$v_i = \log\left(\frac{\sigma_i}{\mu_i}\right), \text{ para } i \text{ em R,G,B,H,S,V} \quad (3.16)$$

3.4.4 Textura

As propriedades de textura descrevem a informação da estrutura do objeto, seus padrões visuais (HARALICK; SHANMUGAM; DINSTEN, 1973). Foram extraídas texturas da matriz de co-ocorrência dos níveis de cinza (GLCM), descritores de Haralick, e da matriz de comprimento dos níveis de cinza (GLRLM).

1. GLCM (T_{1-16})

A matriz GLCM tem o objetivo de capturar a estrutura da imagem considerando a frequência com que pares de pixels com valores específicos aparecem como vizinhos. A vizinhança pode considerar diferentes orientações. Cada uma das quantidades numéricas seguintes foram calculadas em quatro orientações: (0° , 45° , 90° , 135°), resultando em 16 descritores.

$$\text{Contraste} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p(i, j)(i - j)^2 \quad (3.17)$$

$$\text{Correlação} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{ijp(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3.18)$$

$$\text{Energia} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p(i, j)^2 \quad (3.19)$$

$$\text{Homogeneidade} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{p(i, j)}{1 + (i - j)^2} \quad (3.20)$$

2. Haralick (T_{17-29})

Os 13 primeiros descritores de Haralick foram implementados como definidos em Haralick, Shanmugam e Dinstein (1973), usando a biblioteca de visão computacional Mahotas (COELHO, 2013).

3. GLRLM (T_{30-36})

Uma carreira de níveis de cinza é um conjunto de pixels colineares e consecutivos com o mesmo nível de cinza (GALLOWAY, 1975). Da matriz GLRLM, foram extraídas 5 descritores de textura T_{30-34} definidos em Galloway (1975). Em Chu, Sehgal e Greenleaf (1990) foram definidos dois novos descritores T_{35-36} , que usam o valor das carreiras de níveis de cinza e não apenas os comprimentos das carreiras como os descritores de Galloway.

$$H = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j)$$

$$T_{30} = \text{Short Run Emphasis} = \frac{1}{H} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i, j)}{j^2} \quad (3.21)$$

$$T_{31} = \text{Long Run Emphasis} = \frac{1}{H} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} j^2 p(i, j) \quad (3.22)$$

$$T_{32} = \text{Gray Level Nonuniformity} = \frac{1}{H} \sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_r} p(i, j) \right)^2 \quad (3.23)$$

$$T_{33} = \text{Run Level Nonuniformity} = \frac{1}{H} \sum_{j=1}^{N_r} \left(\sum_{i=1}^{N_g} p(i, j) \right)^2 \quad (3.24)$$

$$T_{34} = \text{Run Percentage} = \frac{H}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} jp(i, j)} \quad (3.25)$$

$$T_{35} = \text{LGRE} = \frac{1}{H} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i, j)}{i^2} \quad (3.26)$$

$$T_{36} = HGRE = \frac{1}{H} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} i^2 p(i, j) \quad (3.27)$$

onde a entrada $p(i, j)$ da matriz GLRLM é igual ao número de vezes que a imagem tem uma sequência de pixels com comprimento j e valor i em uma direção específica, N_r é o comprimento da maior carreira e N_g é o número de níveis de cinza.

3.5 Características extraídas com CNN e redução de dimensionalidade

Para este trabalho, baseado em experimentos anteriores com bons resultados (PACHECO; KROHLING, 2020), foi escolhida a arquitetura ResNet-50. A ResNet-50 utilizada neste trabalho foi pré-treinada com o conjunto de dados da ImageNet e retreinada, através de transferência de aprendizado (Em inglês: *transfer learning*) com o conjunto de imagens utilizado, e, fornece em sua saída um vetor de 2048 descritores. Como esse vetor está em um espaço de alta dimensionalidade, é necessário aplicar um algoritmo de redução de dimensionalidade na saída da rede. Para a redução de dimensionalidade do vetor de características foram utilizados dois algoritmos: Análise das componentes principais (PCA) e Projeção e aproximação de variedade uniforme (UMAP) (MCINNES et al., 2018). UMAP é um algoritmo não-linear de redução de dimensionalidade de propósito geral que pode ser usado para redução dos vetores de características e também para visualização, como o t-SNE (MAATEN; HINTON, 2008). O número de dimensões utilizadas é discutido a seguir.

3.6 Fusão das características

Para cada imagem da partição de treino, a ResNet-50 fornece em sua saída um vetor de comprimento 2048. Após a obtenção dos vetores de todas as imagens, é formada uma matriz de dimensões $M \times N$, onde M é o número de imagens e N o tamanho do vetor, 2048 no caso da ResNet-50. Essas características são reduzidas antes da combinação com os descritores calculados, utilizando o PCA ou UMAP. Para obter o número de dimensões resultantes do PCA ou UMAP, foi usado um fator de combinação ϕ_1 , como em Pacheco e Krohling (2020), que determina o percentual ocupado pelas características provenientes da ResNet-50 no vetor formado pela concatenação dessas características com as características calculadas. O número final de descritores da CNN é dado por:

$$F_1 = \left\lceil \frac{\phi_1 \times N_h}{1 - \phi_1} \right\rceil \quad (3.28)$$

onde N_h é sempre 72 e $0 < \phi_1 < 1$.

Como exemplo, um ϕ_1 de 0.7 significa que 70% das características devem vir da ResNet-50 e 30% dos descritores calculados. Neste caso, o processo de redução transformará o vetor inicial de 2048 descritores em um vetor com 168 descritores. A concatenação do vetor resultante com os 72 descritores calculados se torna a entrada para uma rede densa (totalmente conectada) cuja função é combinar as características e fazer a classificação. O mesmo procedimento é usado para definir a razão de combinação entre os descritores provenientes da imagem (CNN e calculados) com as informações clínicas do paciente na próxima camada da rede totalmente conectada. Para definir o número de saídas dessa camada, que equivale a quantas características da imagem serão combinadas com as informações do paciente, foi usada uma versão equivalente da equação 3.28, mas agora com um fator ϕ_2 . O número de características que serão combinadas com os metadados do paciente é dado por:

$$F_2 = \left\lceil \frac{\phi_2 \times N_c}{1 - \phi_2} \right\rceil \quad (3.29)$$

onde $N_c = 24$, é o tamanho do vetor com as informações do paciente. A entrada do classificador (retângulo vermelho na Figura 16) é o vetor resultante da concatenação com as informações clínicas.

4 Experimentos e resultados

4.1 Experimentos

Os experimentos consistiram das seguintes etapas: 1) treino de ajuste da ResNet-50 (*fine-tuning*); 2) extração e redução da dimensionalidade das características obtidas com a ResNet-50 usando PCA e UMAP, com a dimensão final calculada com base no valor de ϕ_1 ; 3) combinação das características calculadas e das informações do paciente; 4) treino da rede totalmente conectada para a fusão das características; 5) classificação da lesão. Foram feitos dois experimentos, o primeiro para discriminar entre Câncer (CAN) vs. Não-Câncer ($\overline{\text{CAN}}$), o segundo para discriminar entre Melanoma vs. Não-Melanoma ($\overline{\text{MEL}}$). A Tabela 3 mostra como os tipos de lesões foram agrupados em classes em cada um dos experimentos. Para cada experimento, foi criado um *baseline* obtido sem combinar as características, i.e, cada tipo de descritor (extraído com CNN, calculado) foi usado separadamente para classificar a lesão. Apenas os dados clínicos do paciente não foram utilizados para a criação do *baseline*, pois são usados como um auxílio para a classificação e não como fonte de informação principal para o diagnóstico ¹.

Diagnóstico	Experimento 1		Experimento 2	
	CAN	$\overline{\text{CAN}}$	MEL	$\overline{\text{MEL}}$
Carcinoma de células basais	X			X
Carcinoma espinocelular	X			X
Melanoma	X		X	
Lentigo maligno	X		X	
Queratose actínica		X		X
Queratose seborréica		X		X
Nevo		X		X
Nº de imagens	456	610	44	1022

Tabela 3 – As sete principais doenças de pele do banco de imagens e como elas foram agrupadas em classes em cada um dos experimentos.

Para o treinamento da ResNet-50 foi usada a técnica de *online augmentation*, que consiste em apresentar para a rede pequenas variações aleatórias das imagens originais do *dataset* a cada época. Foram usadas as seguintes transformações nas imagens: alteração no brilho e contraste, giro horizontal (90°) e vertical (180°) da imagem e uma pequena rotação (entre -7.2° e 7.2°). Para lidar com o desbalanceamento do *dataset*, foram atribuídos pesos a cada classe durante o cálculo da entropia cruzada na função de perda da rede neural. Os pesos da classe i foram calculados usando a seguinte fórmula:

¹ Médicos especialistas usam os dados clínicos e a imagem da lesão para fazer o diagnóstico.

$$W_i = \frac{N}{C \times N_i} \quad (4.1)$$

onde W_i é o peso da classe i , C é o número de classes (2 em todos os experimentos), N é o total de amostras na partição de treino e N_i é o número de amostras da classe i na partição de treino.

Como a ResNet-50 foi pré-treinada no *dataset* da ImageNet, foi preciso fazer um treino de ajuste dos pesos (*fine-tuning*) para as imagens do PAD-UFES-20. A camada de classificação da ResNet-50 original foi removida, e uma nova camada, inicializada aleatoriamente, foi inserida e treinada por 10 épocas, enquanto o restante dos pesos foi mantido fixo. Esse procedimento inicial é feito para evitar que os gradientes elevados que aparecem em virtude da inicialização aleatória dos pesos no início do treinamento, prejudiquem os pesos já ajustados da ResNet-50 (GÉRON, 2019). Após esse treino inicial, a rede inteira foi treinada por mais 100 épocas, utilizando *early stopping* para prevenir o sobreajuste (*overfitting*). Em todos os treinamentos foi usado uma taxa de aprendizagem de 0.001 e o otimizador *stochastic gradient descent* (SGD) (RUDER, 2016). A camada totalmente conectada usada para a combinação das características foi treinada por 100 épocas.

Foram investigadas a combinação de características em 3 contextos: 1) características da CNN e características calculadas; 2) características da CNN e metadados clínicos; 3) todas as características (CNN, calculadas e metadados clínicos). No primeiro caso, foi feita uma busca pelo melhor fator de combinação ϕ_1 , na faixa de 0.5 à 0.9, com incremento de 0.1, para combinar as características da imagem. Foi utilizada a técnica de validação cruzada com 5 partições e escolheu-se o valor de ϕ_1 que obteve o maior valor para a acurácia balanceada no conjunto de validação. Depois de encontrar o melhor valor de ϕ_1 , efetuou-se a avaliação do conjunto de teste. O segundo caso é idêntico ao primeiro, apenas substituindo os descritores calculados pelas informações do paciente. Para o terceiro caso, após a determinação do fator ϕ_1 , foi realizada uma nova validação cruzada para obter o melhor fator de combinação ϕ_2 . A quantidade de descritores utilizados para cada valor dos fatores de combinação estão listados nas Tabelas 4 e 5.

ϕ_1	$f_{ResNet-50}$	$f_{calculadas}$	Total
0.5	72	72	144
0.6	108	72	180
0.7	168	72	240
0.8	288	72	360
0.9	648	72	720

Tabela 4 – Número de características usadas para cada valor de ϕ_1 . $f_{ResNet-50}$ é calculado de acordo com a Equação 3.28.

Para a implementação dos experimentos foi utilizada a linguagem de programação

ϕ_2	f_{imagem}	$f_{paciente}$	Total
0.5	24	24	48
0.6	36	24	60
0.7	56	24	80
0.8	97	24	121
0.9	216	24	240

Tabela 5 – Número de características usadas para cada valor de ϕ_2 . f_{imagem} é calculado de acordo com a Equação 3.29.

Python, juntamente com as bibliotecas Pytorch para a construção e treinamento da CNN, Scikit-learn para algoritmos de classificação e cálculo de métricas, Numpy e Pandas para manipulação dos dados e OpenCV para a implementação das rotinas de extração de características calculadas. O código-fonte dos experimentos pode ser disponibilizado através de solicitação.

4.2 Métricas

Para avaliar os resultados, foram utilizadas 5 métricas conhecidas na literatura: Acurácia (Acc), Acurácia balanceada (Bacc), Precisão (Pr), Recall (Re) e F1, definidas a seguir:

$$Acc = \frac{VP + VN}{VP + FP + FN + VN} \quad (4.2)$$

$$Bacc = \frac{1}{2} \times \left(\frac{VP}{VP + FN} + \frac{VN}{VN + FP} \right) \quad (4.3)$$

$$Pr = \frac{VP}{VP + FP} \quad (4.4)$$

$$Re = \frac{VP}{VP + FN} \quad (4.5)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4.6)$$

onde, VP (*verdadeiros positivos*), é o número de amostras classificadas corretamente como positivas, VN (*verdadeiros negativos*), é o número de amostras classificadas corretamente como negativas, FP (*falsos positivos*), é a quantidade de amostras classificadas equivocadamente como positivas e, FN (*falsos negativos*), é o número de amostras classificadas equivocadamente como negativas. Sendo a acurácia balanceada, a principal métrica utilizada, pois ela leva em consideração o desbalanceamento entre as classes, calculando a média entre os acertos nas classes positivas e negativas.

4.3 Resultados

Nessa seção, são apresentados os resultados da validação cruzada para a escolha dos melhores fatores de combinação e os resultados no conjunto de teste, utilizando os fatores de combinação selecionados.

4.3.1 Câncer vs. Não-câncer

As Tabelas 6 e 7 listam os resultados da validação cruzada para o caso da combinação entre os descritores da CNN e os descritores calculados, utilizando PCA e UMAP como os métodos de redução de dimensionalidade, respectivamente. De acordo com a Tabela 6, o melhor valor de ϕ_1 usando PCA é 0.5. Embora o PCA alcançou os melhores resultados de acurácia balanceada e *recall*, os resultados obtidos com o uso do algoritmo UMAP são mais estáveis para os diferentes valores de ϕ_1 . Depois de encontrar o melhor valor para ϕ_1 , a rede totalmente conectada, para a combinação dos descritores, foi treinada, testando uma faixa de valores para ϕ_2 . As Tabelas 8 e 9 listam os resultados para os diferentes valores de ϕ_2 , enquanto ϕ_1 permanece em seu melhor valor, usando PCA e UMAP, respectivamente, para redução de dimensionalidade. Depois de encontrar os melhores valores para ϕ_1 e ϕ_2 , a partição de teste é avaliada. Os resultados estão apresentados na Tabela 10 e mostram uma notável melhora nas métricas, o que indica que a combinação dos três tipos de descritores pode melhorar a classificação de câncer de pele. A partir da Tabela 10, é possível concluir que apenas a combinação de características da CNN e características calculadas não melhorou os resultados comparado ao uso somente da CNN, enquanto o uso dos três tipos de características melhorou consistentemente os resultados. A Figura 17 mostra as matrizes de confusão obtidas nos experimentos. É possível observar um incremento de 66,07% para 80,36% na identificação de amostras positivas de câncer utilizando os 3 tipos de características quando comparado com apenas a ResNet-50, enquanto a taxa de identificação de amostras negativas permaneceu constante, obtendo assim, aumento significativo no *Recall*. A curva ROC na Figura 18 e a Tabela 10 indicam que, a combinação dos três tipos de descritores, resultou em uma melhor precisão e recall, mantendo a área sob a curva ROC similar. A Figura 18 também mostra que o uso das características calculadas em conjunto com as da CNN resultou em um decréscimo na performance do classificador quando comparado com o uso apenas da CNN.

ϕ_1	Acc %	Bacc %	Pr %	Re %	F1 %
0.5	82.7 ± 3.6	82.4 ± 3.9	78.7 ± 3.7	80.5 ± 6.4	79.5 ± 4.5
0.6	80.3 ± 3.0	79.9 ± 3.1	76.1 ± 4.5	77.3 ± 4.9	76.6 ± 3.6
0.7	78.0 ± 3.5	77.8 ± 3.7	72.2 ± 4.1	77.0 ± 6.1	74.4 ± 4.3
0.8	78.5 ± 3.3	77.9 ± 3.3	74.3 ± 4.3	74.4 ± 3.8	74.3 ± 3.8
0.9	76.8 ± 4.4	75.6 ± 4.8	73.8 ± 4.8	68.5 ± 7.5	71.0 ± 5.9

Tabela 6 – Resultados da validação cruzada para a combinação de características da CNN com as calculadas no experimento Câncer vs. Não-Câncer e PCA para a redução de dimensionalidade.

ϕ_1	Acc %	Bacc %	Pr %	Re %	F1 %
0.5	82.4 ± 1.8	81.1 ± 2.2	82.4 ± 3.4	73.8 ± 5.64	77.7 ± 2.8
0.6	82.1 ± 2.7	80.6 ± 3.4	83.1 ± 2.1	71.7 ± 7.4	76.8 ± 4.4
0.7	82.4 ± 2.2	81.1 ± 2.5	82.4 ± 3.8	73.8 ± 6.0	77.7 ± 3.0
0.8	82.4 ± 2.7	81.2 ± 2.5	82.4 ± 6.2	74.3 ± 4.6	77.9 ± 2.9
0.9	81.7 ± 2.3	80.4 ± 2.3	82.0 ± 4.0	72.3 ± 3.2	76.8 ± 2.8

Tabela 7 – Resultados da validação cruzada para a combinação de características da CNN com as calculadas no experimento Câncer vs. Não-Câncer e UMAP para a redução de dimensionalidade.

ϕ_1	ϕ_2	Acc %	Bacc %	Pr %	Re %	F1 %
0.5	0.5	85.9 ± 2.6	85.4 ± 2.7	83.4 ± 3.8	82.8 ± 4.3	83.0 ± 3.2
0.5	0.6	84.4 ± 0.8	84.3 ± 1.0	80.0 ± 1.6	83.7 ± 3.4	81.7 ± 1.3
0.5	0.7	83.4 ± 3.1	83.3 ± 3.2	78.9 ± 3.9	82.5 ± 4.4	80.6 ± 3.6
0.5	0.8	85.8 ± 3.0	85.7 ± 3.2	81.6 ± 3.6	85.1 ± 4.9	83.3 ± 3.6
0.5	0.9	84.4 ± 3.3	84.1 ± 3.4	81.1 ± 4.3	81.9 ± 4.6	81.4 ± 3.9

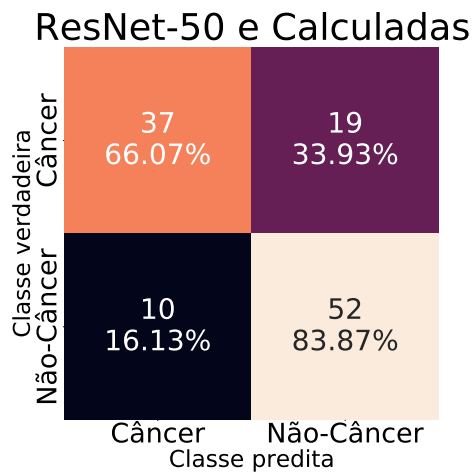
Tabela 8 – Resultados da validação cruzada para a combinação dos três tipos de descritores no experimento Câncer vs. Não-Câncer utilizando PCA.

ϕ_1	ϕ_2	Acc %	Bacc %	Pr %	Re %	F1 %
0.8	0.5	85.6 ± 3.0	84.6 ± 3.4	86.1 ± 3.0	78.1 ± 5.6	81.9 ± 4.2
0.8	0.6	86.5 ± 2.0	85.9 ± 2.1	84.9 ± 3.7	82.5 ± 4.4	83.6 ± 2.4
0.8	0.7	86.8 ± 1.8	85.9 ± 2.0	87.1 ± 2.6	80.5 ± 3.6	83.6 ± 2.4
0.8	0.8	85.3 ± 1.0	84.3 ± 1.0	85.1 ± 2.8	78.7 ± 2.7	81.7 ± 1.2
0.8	0.9	86.1 ± 1.5	85.3 ± 1.5	85.6 ± 2.7	80.5 ± 3.1	82.9 ± 1.9

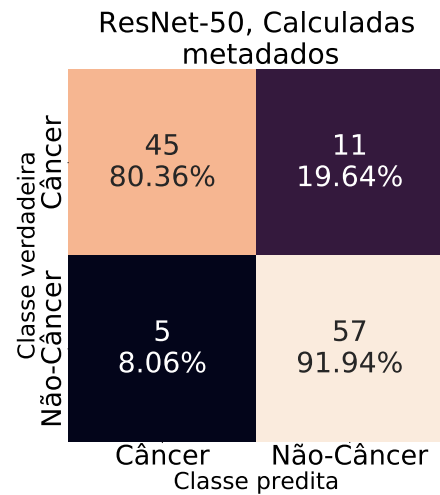
Tabela 9 – Resultados da validação cruzada para a combinação dos três tipos de descritores no experimento Câncer vs. Não-Câncer utilizando UMAP.

Descritores	Redução	ϕ_1	ϕ_2	Acc %	Bacc %	Pr %	Re %	F1 %
ResNet-50 + Handcrafted	UMAP	0.8	-	78.8	77.9	91.8	60.7	73.1
ResNet-50 + Handcrafted	PCA	0.5	-	75.4	74.9	78.7	66.0	71.8
ResNet-50 + Handcrafted + Clinical	UMAP	0.8	0.7	81.3	80.7	90.4	67.8	77.5
ResNet-50 + Handcrafted + Clinical	PCA	0.5	0.8	86.4	86.1	90.0	80.3	84.9
ResNet-50 + Clinical	PCA	-	-	82.2	81.8	85.7	75.0	80.0
ResNet-50 + Clinical	UMAP	-	-	78.8	78.1	87.8	64.2	74.2
Only Handcrafted (SVM)	PCA	-	-	74.5	74.2	76.0	67.8	71.6
Only ResNet-50	-	-	-	79.6	79.0	88.0	66.0	75.5

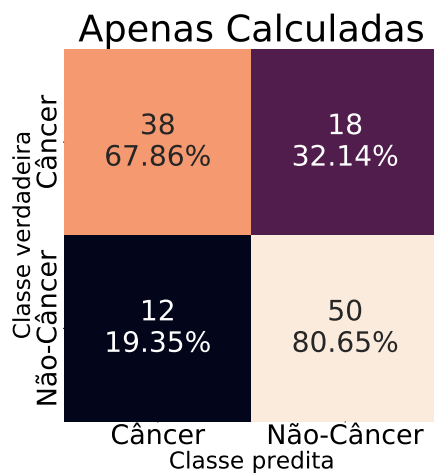
Tabela 10 – Resultados do experimento Câncer vs. Não-Câncer nos dados de teste.



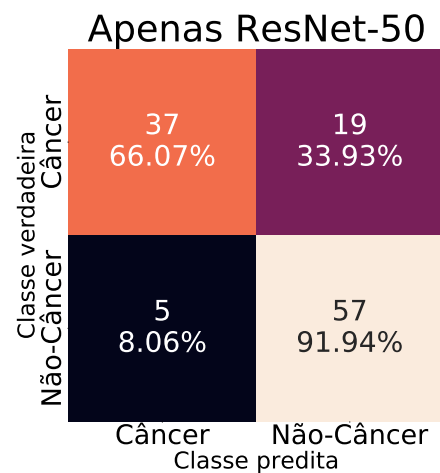
(a) ResNet-50 e descritores calculados (PCA).



(b) Os 3 tipos de descritores (PCA).



(c) Apenas descritores calculados.



(d) Apenas ResNet-50.

Figura 17 – Matrizes de confusão do experimento Câncer vs. Não-Câncer, mostrando a melhora no uso dos três tipos de descritores.

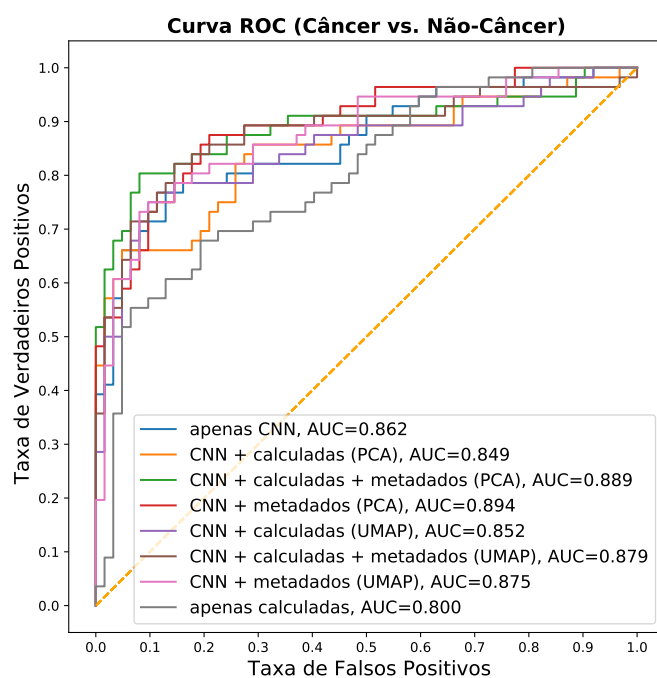


Figura 18 – Curva ROC do experimento Câncer vs. Não-Câncer.

4.3.2 Melanoma vs. Não-Melanoma

As Tabelas 11 e 12 mostram os resultados da avaliação do fator de combinação ϕ_1 do experimento Melanoma vs. Não-Melanoma. Novamente, o melhor valor para ϕ_1 foi 0.5 usando PCA. Os resultados inferiores obtidos quando ϕ_1 é 0.9 podem ser explicados pelo fato de que o uso de mais componentes principais introduzem maior variabilidade e ruído no classificador. As Tabelas 13 e 14 contêm os resultados da determinação do melhor valor para ϕ_2 . A Tabela 15 mostra os resultados finais no conjunto de teste utilizando os melhores valores encontrados de ϕ_1 e ϕ_2 . Os resultados da Tabela 15 apresentam o mesmo padrão dos experimentos anteriores. A combinação dos descritores calculados com os da CNN resultou em métricas inferiores quando comparado com o uso somente da ResNet-50. As matrizes de confusão na Figura 19 mostram que o uso dos três tipos de descritores contribuiu para uma melhor discriminação do melanoma. É possível observar que ao utilizar os 3 tipos de características houve um decréscimo na taxa de identificação de amostras negativas de melanoma quando comparado com apenas o uso das características extraídas da ResNet-50. A Figura 20 mostra que a combinação dos três tipos de descritores fornece o maior valor para a métrica AUC, enquanto usar apenas as características extraídas com a CNN em conjunto com as calculadas resultou em um pior resultado. Como no experimento anterior, o melhor resultado foi obtido com a combinação dos três tipos de descritores, seguido do uso somente da CNN. As matrizes de confusão na Figura 19 mostram que o uso de todos os descritores pode melhorar a detecção do melanoma em troca de um aumento no número de falsos positivos. Usar somente as características calculadas fornece resultados inferiores de classificação, o que pode ser atribuído ao fato de que imagens clínicas não possuem o mesmo nível de detalhes que imagens dermatoscópicas possuem.

ϕ_1	Acc %	Bacc %	Pr %	Re %	F1 %
0.5	93.3 ± 1.3	60.6 ± 8.5	12.6 ± 7.7	26.0 ± 17.4	16.8 ± 10.6
0.6	94.1 ± 1.6	57.1 ± 8.2	13.5 ± 12.0	18.0 ± 15.6	15.3 ± 13.3
0.7	92.4 ± 1.4	53.3 ± 8.1	6.5 ± 9.6	12.0 ± 16.0	8.3 ± 11.0
0.8	94.9 ± 1.7	55.1 ± 5.3	10.8 ± 9.1	13.0 ± 10.7	11.6 ± 9.5
0.9	95.7 ± 0.7	49.2 ± 0.2	0 ± 0	0 ± 0	0 ± 0

Tabela 11 – Resultados da validação cruzada usando a combinação das características da ResNet-50 e calculadas, no experimento Melanoma vs. Não-Melanoma utilizando PCA para a redução da dimensionalidade.

ϕ_1	Acc %	Bacc %	Pr %	Re %	F1 %
0.5	91.0 \pm 1.1	63.8 \pm 11.7	10.5 \pm 5.9	35.0 \pm 24.8	16.0 \pm 9.5
0.6	88.2 \pm 3.0	61.3 \pm 12.2	8.3 \pm 5.1	33.0 \pm 26.7	13.0 \pm 8.1
0.7	89.0 \pm 2.8	62.8 \pm 9.6	9.0 \pm 5.7	35.0 \pm 20.9	14.2 \pm 8.8
0.8	90.4 \pm 1.7	54.7 \pm 8.1	6.7 \pm 6.3	17.0 \pm 15.3	9.5 \pm 8.8
0.9	88.2 \pm 1.8	66.2 \pm 16.8	8.7 \pm 7.2	43.0 \pm 35.7	14.5 \pm 12.1

Tabela 12 – Resultados da validação cruzada usando a combinação das características da ResNet-50 e calculadas, no experimento Melanoma vs. Não-Melanoma utilizando UMAP para a redução da dimensionalidade.

ϕ_1	ϕ_2	Acc %	Bacc %	Pr %	Re %	F1 %
0.5	0.5	93.5 \pm 1.6	62.1 \pm 9.8	15.1 \pm 10.7	29.0 \pm 20.0	19.5 \pm 13.6
0.5	0.6	93.4 \pm 2.0	59.2 \pm 10.7	13.3 \pm 11.3	23.0 \pm 20.3	16.8 \pm 14.4
0.5	0.7	93.8 \pm 3.1	65.7 \pm 11.5	19.8 \pm 15.3	36.0 \pm 22.4	24.9 \pm 17.8
0.5	0.8	92.9 \pm 2.5	64.8 \pm 16.1	17.7 \pm 15.0	35.0 \pm 30.6	23.5 \pm 20.0
0.5	0.9	91.8 \pm 1.2	56.0 \pm 7.9	7.5 \pm 6.9	18.0 \pm 15.6	10.5 \pm 9.5

Tabela 13 – Resultados da validação cruzada para a combinação dos três tipos de descritores no experimento Melanoma vs. Não-Melanoma utilizando PCA.

ϕ_1	ϕ_2	Acc %	Bacc %	Pr %	Re %	F1%
0.9	0.5	90.9 \pm 4.2	79.3 \pm 10.7	20.5 \pm 7.5	67.0 \pm 19.8	31.0 \pm 11.0
0.9	0.6	88.4 \pm 5.6	68.3 \pm 11.2	14.5 \pm 9.6	47.0 \pm 22.2	21.2 \pm 12.5
0.9	0.7	90.0 \pm 0.9	74.5 \pm 16.1	14.0 \pm 7.7	58.0 \pm 33.7	22.4 \pm 12.3
0.9	0.8	88.9 \pm 5.0	69.5 \pm 10.0	15.6 \pm 10.1	49.0 \pm 18.5	22.8 \pm 13.0
0.9	0.9	87.2 \pm 4.8	70.6 \pm 7.4	13.1 \pm 6.3	53.0 \pm 13.2	20.6 \pm 8.7

Tabela 14 – Resultados da validação cruzada para a combinação dos três tipos de descritores no experimento Melanoma vs. Não-Melanoma utilizando UMAP.

Descritores	Redução	ϕ_1	ϕ_2	Acc %	Bacc %	Pr %	Re %	F1 %
ResNet-50 + Handcrafted	UMAP	0.9	-	92.3	83.5	57.5	72.7	64.0
ResNet-50 + Handcrafted	PCA	0.5	-	92.3	71.3	62.5	45.4	52.6
ResNet-50 + Handcrafted + Clinical	UMAP	0.9	0.5	93.2	88.1	60.0	81.8	69.2
ResNet-50 + Handcrafted + Clinical	PCA	0.5	0.7	91.5	87.1	52.9	81.8	64.2
ResNet-50 + Clinical	PCA	-	-	94.0	80.4	70.0	63.6	66.6
ResNet-50 + Clinical	UMAP	-	-	94.0	84.4	66.6	72.7	69.5
Only Handcrafted (SVM)	PCA	-	-	89.8	57.6	40.0	18.1	25.0
Only ResNet-50	-	-	-	94.9	84.9	72.7	72.7	72.7

Tabela 15 – Resultados do experimento Melanoma vs. Não-Melanoma nos dados de teste.

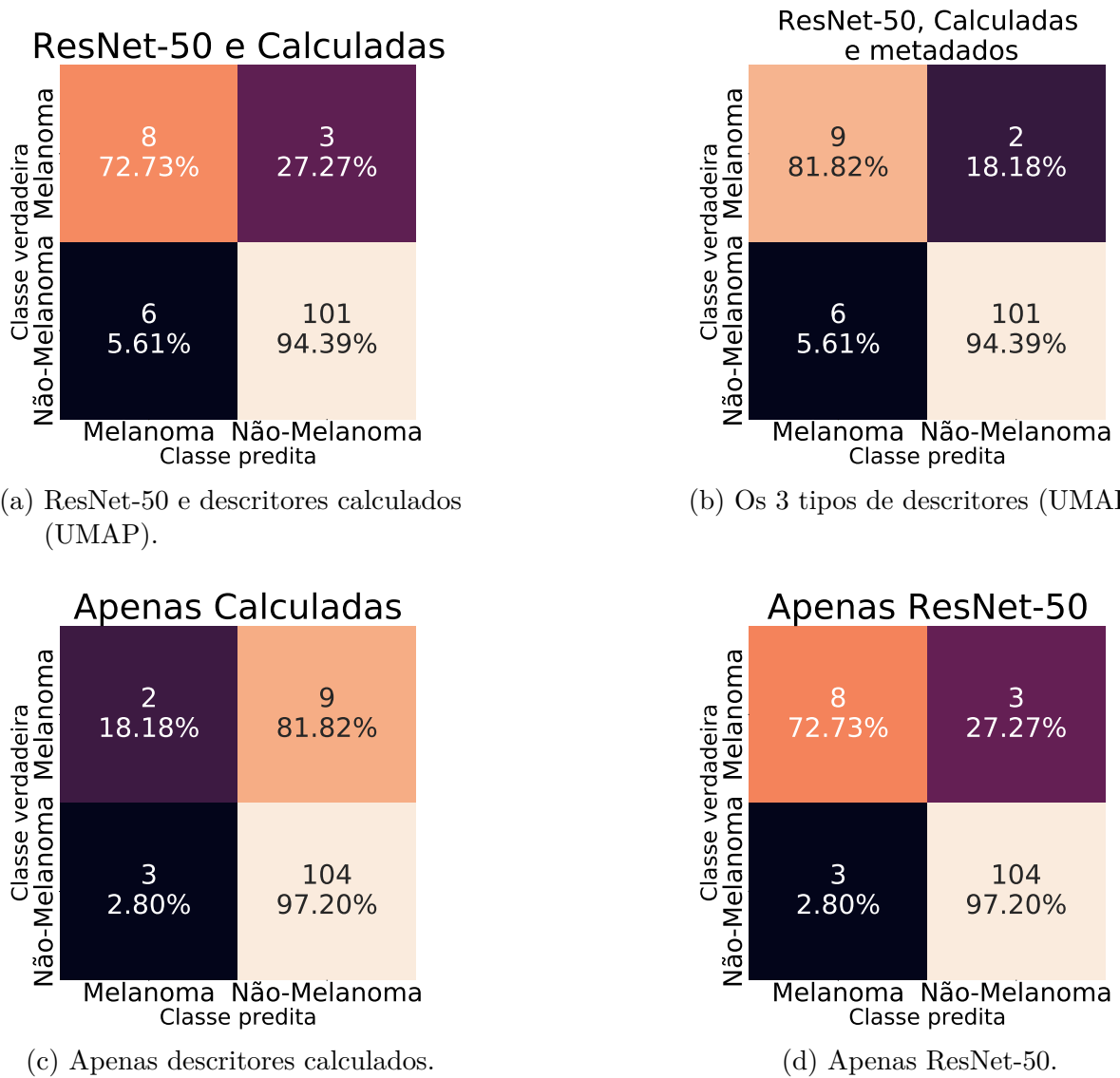


Figura 19 – Matrizes de confusão do experimento Melanoma vs. Não-Melanoma.

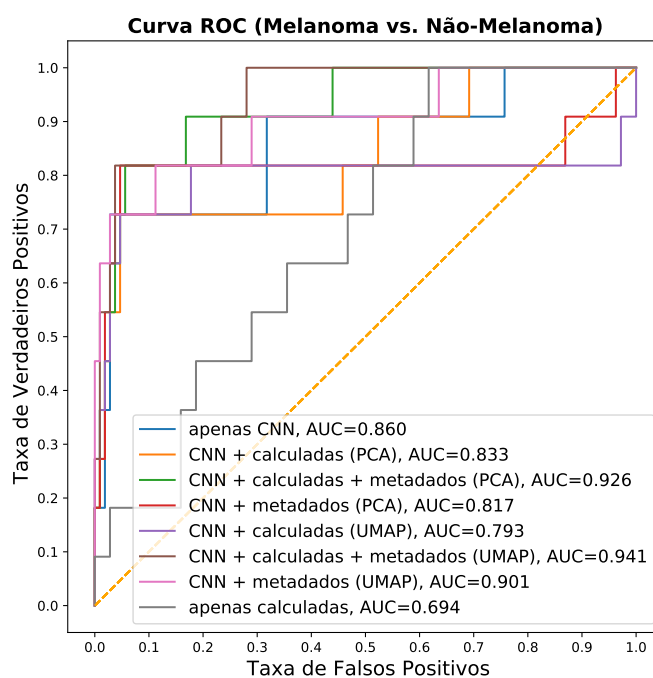


Figura 20 – Curva ROC do experimento Melanoma vs. Não-Melanoma.

5 Conclusão

Neste trabalho foi proposto uma abordagem para a combinação de três tipos de características (extraídas da CNN, calculadas e metadados) com o propósito de classificar imagens clínicas de lesões de pele utilizando redes neurais. A abordagem foi avaliada no banco de imagens PAD-UFES-20 e os resultados indicam que a combinação dos três tipos de características fornecem os melhores resultados de classificação para ambos os casos: Câncer vs. Não-Câncer e Melanoma vs. Não-Melanoma. A abordagem proposta usando os metadados clínicos em conjunto com as características extraídas via CNN e os descritores calculados, aumentaram a acurácia balanceada de 79.0% para 86.1% para a classificação de câncer e de 84.9% para 88.1% para a classificação de melanoma, ao mesmo tempo, melhorando o recall (sensitividade), de 66% para 80.3% no caso do câncer e de 72.7% para 81.8% para o melanoma. A combinação de apenas os descritores calculados com os descritores extraídos via CNN, não resultou em melhorias, o que reforça a importância do uso dos metadados clínicos como complemento às imagens clínicas, já que esse tipo de imagem não apresenta o mesmo nível de informação visual que as imagens dermatoscópicas. Não foi observada diferença significativa com relação aos métodos de redução de dimensionalidade, exceto que o PCA consistentemente utilizou menos componentes para obter resultados semelhantes aos obtidos pelo algoritmo UMAP. Embora os experimentos indiquem que houve melhoria ao utilizar os 3 tipos de características em conjunto, o sistema ainda não pode ser usado na prática pois os resultados precisam ser avaliados em um conjunto maior de imagens e o problema da segmentação precisa ser resolvido para que a extração das características calculadas possa ser feito automaticamente sem o auxílio de máscaras feitas previamente. Problemas de classificação de padrões se beneficiam da quantidade de dados disponíveis, neste trabalho, foi possível notar que embora as técnicas utilizadas indicaram melhorias na classificação, o grande desbalanceamento entre os tipos de lesões pode influenciar a robustez dos resultados e deve ser priorizado em trabalhos futuros. Como neste trabalho foram utilizadas técnicas de aprendizado supervisionado, uma possibilidade de pesquisa futura seria o uso de algoritmos e arquiteturas de redes neurais baseadas em aprendizado não-supervisionado, como algoritmos de agrupamento e *autoencoders*, assim, o sistema seria independente dos rótulos das lesões. Outra possibilidade seria usar a combinação de CNNs e a arquitetura de *Transformers*, pois enquanto a CNN aprende características baseadas em textura, a rede *Transformer* aprende características baseadas na forma dos objetos (WEI et al., 2022), dessa forma, o sistema dependeria menos da extração das características calculadas e do processo de segmentação.

Referências

- ALFED, N.; KHELIFI, F. Bagged textural and color features for melanoma skin cancer detection in dermoscopic and standard images. *Expert Systems with Applications*, v. 90, p. 101 – 110, 2017.
- ALI, K. et al. Multiclass skin cancer classification using efficientnets – a first step towards preventing skin cancer. *Neuroscience Informatics*, v. 2, n. 4, p. 100034, 2022. Artificial Intelligence in Brain Informatics.
- ARGENZIANO, G. et al. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Archives of Dermatology*, v. 134, n. 12, p. 1563–1570, 12 1998.
- ASIRI, N. et al. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. *Artificial Intelligence in Medicine*, v. 99, p. 101701, 2019.
- BARATA, C.; CELEBI, M. E.; MARQUES, J. S. A survey of feature extraction in dermoscopy image analysis of skin cancer. *IEEE Journal of Biomedical and Health Informatics*, v. 23, n. 3, p. 1096–1109, 2019.
- BARATA, C. et al. Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Systems Journal*, v. 8, n. 3, p. 965–979, 2014.
- BHUIYAN, M. A. H.; AZAD, I.; UDDIN, K. Image processing for skin cancer features extraction. *International Journal of Scientific & Engineering Research*, v. 4, n. 2, p. 1–6, 2013.
- BINDER, M. et al. Epiluminescence microscopy: A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Archives of Dermatology*, v. 131, n. 3, p. 286–291, 03 1995.
- BISHOP, C. *Neural networks for pattern recognition*. [S.l.]: Oxford University Press, USA, 1995.
- BISHOP, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. 1. ed. [S.l.]: Springer, 2007. ISBN 0387310738.
- BRINKER, T. J.; HEKLER, A.; ENK, A. H. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer*, v. 113, p. 47–54, 2019.
- CHATTERJEE, S. et al. Extraction of features from cross correlation in space and frequency domains for classification of skin lesions. *Biomedical Signal Processing and Control*, v. 53, p. 101581, 2019.
- CHERIFI, H. Interpretable aide diagnosis system for melanoma recognition. *Journal of Bioengineering & Biomedical Science*, v. 04, 01 2014.

- CHU, A.; SEHGAL, C.; GREENLEAF, J. Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*, v. 11, n. 6, p. 415–419, 1990.
- CODELLA, N. et al. Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. In: *Machine Learning in Medical Imaging*. [S.l.]: Springer International Publishing, 2015. p. 118–126.
- COELHO, L. P. Mahotas: Open source software for scriptable computer vision. *Journal of Open Research Software* 1(1):e3, 2013.
- De Angelo, G. G.; PACHECO, A. G. C.; KROHLING, R. A. Skin lesion segmentation using deep learning for images acquired from smartphones. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2019. p. 1–8.
- DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: IEEE. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.], 2009. p. 248–255.
- DEVIHA, V. A study on skin cancer using fractals. *International Journal of Mathematics and its Applications*, v. 4, p. 121–126, 2016.
- ESGARIO, J. G. M.; KROHLING, R. A. Beyond visual image: Automated diagnosis of pigmented skin lesions combining clinical image features with patient data. *arXiv e-prints*, arXiv:2201.10650, 2022.
- ESTEVA, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, v. 542, n. 7639, p. 115–118, 2017.
- FERRARA, G. et al. The influence of clinical information in the histopathologic diagnosis of melanocytic skin neoplasms. *PloS one*, v. 4, n. 4, p. e5375–e5375, 2009.
- GALLOWAY, M. M. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, v. 4, n. 2, p. 172–179, 1975.
- GÉRON, A. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. Second. [S.l.]: O’Reilly Media, 2019.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016.
- HAMEED, N. et al. Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques. *Expert Systems with Applications*, v. 141, p. 112961, 2020.
- HAMIDINEKOO, A. et al. Deep learning in mammography and breast histology, an overview and future trends. *Medical Image Analysis*, v. 47, p. 45–67, 2018.
- HARALICK, R.; SHANMUGAM, K.; DINSTEN, I. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3, n. 6, p. 610–621, 1973.
- HAYKIN, S. S. *Neural networks and learning machines*. Third edition. Upper Saddle River, NJ: Pearson Education, 2009.
- HE, K. et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778.

- HELFAND, M. et al. Screening for skin cancer. *American Journal of Preventive Medicine*, v. 20, n. 3, p. 47–58, Apr 2001.
- HORN, B. K. P. *Robot Vision*. [S.l.]: The MIT Press, Cambridge, Massachusetts., 1986.
- HOWARD, A. G. et al. *MobileNets: efficient convolutional neural networks for mobile vision applications*. [S.l.]: arXiv, 2017.
- HUANG, G. et al. Densely connected convolutional networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 2261–2269.
- HUBEL, D. H.; WIESEL, T. N. Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, v. 148, p. 574–591, 1959.
- JAWOREK-KORJAKOWSKA, J.; TADEUSIEWICZ, R. Determination of border irregularity in dermoscopic color images of pigmented skin lesions. In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. [S.l.: s.n.], 2014. p. 6459–6462.
- KASPRZAK, J.; XU, Y. Diagnosis and management of lentigo maligna: a review. *Drugs in Context*, Just Medical Media Limited, v. 4, p. 212281, May 2015.
- KAWAHARA, J.; BENTAIEB, A.; HAMARNEH, G. Deep features to classify skin lesions. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. [S.l.: s.n.], 2016. p. 1397–1400.
- KENNEDY, J.; EBERHART, R. Particle swarm optimization. In: *Proceedings of ICNN'95 - International Conference on Neural Networks*. [S.l.: s.n.], 1995. v. 4, p. 1942–1948 vol.4.
- KITTLER, H. et al. Diagnostic accuracy of dermoscopy. *The Lancet Oncology*, v. 3, n. 3, p. 159–165, Mar 2002.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, v. 2, p. 1097–1105, 2012.
- KROHLING, R. A. Comunicado privado. PPGI/UFES, 2020.
- LECUN, Y. et al. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, v. 1, n. 4, p. 541–551, 12 1989.
- LI, W. et al. Fusing metadata and dermoscopy images for skin disease diagnosis. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. [S.l.: s.n.], 2020. p. 1996–2000.
- LINARES, M. A.; ZAKARIA, A.; NIZRAN, P. Skin cancer. *Primary Care: Clinics in Office Practice*, v. 42, n. 4, p. 645 – 659, 2015. Primary Care Dermatology.
- LIU, X. et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*, v. 1, n. 6, p. e271–e297, 2019.
- MAATEN, L. van der; HINTON, G. Visualizing data using t-sne. *Journal of Machine Learning Research*, v. 9, n. 86, p. 2579–2605, 2008.

- MAJTNER, T.; YILDIRIM-YAYILGAN, S.; HARDEBERG, J. Combining deep learning and hand-crafted features for skin lesion classification. In: *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. [S.l.: s.n.], 2016. p. 1–6.
- MCCULLOCH, W.; PITTS, W. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, v. 5, p. 127–147, 1943.
- MCINNES, L. et al. UMAP: Uniform manifold approximation and projection. *Journal of Open Source Software*, The Open Journal, v. 3, n. 29, p. 861, 2018.
- NACHBAR, F. et al. The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology*, v. 30, n. 4, p. 551–559, Apr 1994.
- PACHECO, A.; KROHLING, R. A. The impact of patient clinical information on automated skin cancer detection. *Computers in Biology and Medicine*, v. 116, p. 103545, 2020.
- PACHECO, A. G. et al. PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data in Brief*, v. 32, p. 106221, 2020.
- RAJPARA, S. et al. Systematic review of dermoscopy and digital dermoscopy/artificial intelligence for the diagnosis of melanoma. *British Journal of Dermatology*, v. 161, n. 3, p. 591–604, 2009.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, n. 6, p. 386–408, 1958.
- RUDER, S. *An overview of gradient descent optimization algorithms*. [S.l.]: arXiv, 2016.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning Representations by Back-propagating Errors. *Nature*, v. 323, n. 6088, p. 533–536, 1986.
- SHEN, D.; WU, G.; SUK, H.-I. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, v. 19, p. 221–248, Jun 2017. 28301734[pmid].
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations*. [S.l.: s.n.], 2015. p. 1–14.
- SMAOUI, N.; BESSASSI, S. A developed system for melanoma diagnosis. *International Journal of Computer Vision and Signal Processing*, v. 3, n. 1, p. 10–17, 2013.
- SZEGEDY, C. et al. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. p. 1–9.
- TAN, M.; LE, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th International Conference on Machine Learning*. [S.l.: s.n.], 2019. v. 97, p. 6105–6114.
- TAN, T. Y.; ZHANG, L.; LIM, C. P. Intelligent skin cancer diagnosis using improved particle swarm optimization and deep learning models. *Applied Soft Computing*, v. 84, p. 105725, 2019.

WEI, H.-P. et al. A comparative study of cnn- and transformer-based visual style transfer. *Journal of Computer Science and Technology*, v. 37, n. 3, p. 601–614, 2022.

World Cancer Research Fund. *Skin cancer statistics / World Cancer Research Fund International*. 2020. <<https://www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics>>. [Online; accessed 27-July-2020].

World Health Organization. 2020. <<https://www.who.int/uv/faq/skincancer/en/index1.html>>. [Online; accessed 27-July-2020].

WU, J. et al. A multi-input CNNs with attention for skin lesion classification. In: *2020 IEEE International Conference on Smart Cloud (SmartCloud)*. Los Alamitos, CA, USA: [s.n.], 2020. p. 78–83.

YU, Z. et al. Convolutional descriptors aggregation via cross-net for skin lesion recognition. *Applied Soft Computing*, v. 92, p. 106281, 2020.

ZHANG, J. et al. Attention residual learning for skin lesion classification. *IEEE Transactions on Medical Imaging*, v. 38, n. 9, p. 2092–2103, 2019.