



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de São José do Rio Preto

Tiago Pinheiro Camargo

**Uso de Mapas de Ativação de Classe em CNN: Um Estudo
Envolvendo a Classificação de Imagens H&E**

São José do Rio Preto
2022

Tiago Pinheiro Camargo

Uso de Mapas de Ativação de Classe em CNN: Um Estudo Envolvendo a Classificação de Imagens H&E

Trabalho de Conclusão de Curso (TCC) apresentado como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação, junto ao Curso de Bacharelado em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Orientador:

Prof. Dr. Leandro Alves Neves

**São José do Rio Preto
2022**

C172u	Camargo, Tiago Pinheiro Uso de mapas de ativação de classe em CNN: um estudo envolvendo a classificação de imagens H&E / Tiago Pinheiro Camargo. -- São José do Rio Preto, 2023 30 p. Trabalho de conclusão de curso (Bacharelado - Ciência da Computação) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto Orientador: Leandro Alves Neves 1. Processamento de imagens. 2. Redes neurais (Computação). 3. Reconhecimento de padrões. 4. Cólon (Anatomia) Câncer. I. Título.
-------	--

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Tiago Pinheiro Camargo

Uso de mapas de Ativação de Classe em CNN: Um Estudo Envolvendo a Classificação de Imagens H&E

Trabalho de Conclusão de Curso (TCC) apresentado como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação, junto ao Curso de Bacharelado em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Comissão Examinadora:

Prof. Dr. Leandro Alves Neves
UNESP – Câmpus de São José do Rio Preto
Orientador

Profa. Dra. Adriana Barbosa Santos
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Rodrigo Capobianco Guido
UNESP – Câmpus de São José do Rio Preto

**São José do Rio Preto
2022**

AGRADECIMENTOS

Primeiramente, eu agradeço a minha família, em especial meus pais e irmãos, que sempre me apoiaram e me aconselharam em minhas decisões, me guiando nos momentos que mais precisei e servindo como inspirações e exemplos para mim.

Agradeço a meus amigos de turma que trilharam comigo essa jornada por todos os momentos e memórias que dividimos, tornando este caminho mais alegre e prazeroso, mesmo nos momentos mais difíceis.

Agradeço a meus amigos da minha cidade natal por permanecerem ao meu lado me apoiando e auxiliando quando precisei ao longo dos anos, deixando a minha vida mais divertida e alegre.

Por fim, agradeço a todos os professores que dividiram seus conhecimentos comigo e com meus colegas de forma exemplar durante meu período na universidade. Em especial, gostaria também de agradecer ao meu orientador Prof. Dr. Leandro Alves Neves, por ter me guiado e pela compreensão e ensinamentos passados durante o tempo que trabalhamos juntos.

Resumo

Sistemas computacionais de apoio a decisão têm sido amplamente explorados para o reconhecimento de padrões histopatológicos, contribuindo com diagnósticos e prognósticos mais precisos. Neste trabalho, imagens histológicas foram analisadas por meio de uma Rede Neural Convolucional (ResNet-50). As características obtidas da última camada convolucional foram utilizadas para representar mapas de ativação de classe. Cada mapa indicou as regiões de interesse mais exploradas no processo de classificação, proporcionando representações para explicar as predições. Adicionalmente, métricas de desempenhos e um índice de relevância foram extraídos para cada amostra, a fim de confirmar e/ou confrontar as taxas de acurácias fornecidas pela rede. Neste contexto, as informações obtidas são úteis para verificar possíveis padrões em diferentes tipos de imagens histológicas, contribuindo para o aprimoramento de estratégias computacionais ou de processos de avaliação por especialistas.

Palavras-chave: Imagens histológicas, reconhecimento de padrões, mapas de ativação de classe, redes neurais convolucionais.

Abstract

Computer-aided diagnoses have been widely explored for the recognition of histopathological patterns, contributing to more accurate decisions and prognoses. In this work, histological images were analyzed by using a Convolutional Neural Network (ResNet-50). The obtained features from the last convolutional layer were applied to represent class activation mapping. Each map indicated the most explored regions of interest in the classification process, providing representations to explain the predictions. In addition, metrics and a relevance index were defined for each classification in order to confirm and/or compare the accuracy rates provided by the network. In this context, the obtained information is useful to verify patterns in different types of histological images, contributing to the improvement of computational strategies or evaluation processes by specialists.

Keywords: Histological images, pattern recognition, class activation mapping, convolutional neural networks.

Lista de Figuras

2.1	Ilustração de uma arquitetura de uma CNN, envolvendo as camadas convolucionais, <i>pooling</i> e totalmente conectadas (classificador neural).	4
2.2	Ilustração da estratégia de <i>shortcut path</i> ou <i>skip connections</i>	6
2.3	Ilustração de um bloco residual na CNN ResNet-50.	7
2.4	Ilustração geral da arquitetura ResNet-50.	7
2.5	Esquema básico de transferência de aprendizado.	8
2.6	Ilustração de um mapa de ativação de classe.	9
2.7	Ilustração do mapeamento de classe de ativação.	9
3.1	Ilustração do modelo proposto contendo as etapas: 1) Pré-processamento; e 2) mapas de classe de ativação.	12
3.2	Exemplos de imagens H&E: (a) e (b) da base CR; (c) e (d) da base OED; (e) e (f) da base UCSB.	16
4.1	Ilustração da evolução de aprendizado e classificação da ResNet-50 nas imagens H&E: (a) CR; (b) OED; (c) UCSB.	19
4.2	Exemplos de imagens histológicas H&E e mapas de ativação correspondentes para explicar as classificações via ResNet-50, envolvendo os conjuntos UCSB ((a) e (b)), OED ((c) e (d)), e CR ((f) e (g)). . . .	21

Lista de Tabelas

3.1	Um resumo dos conjuntos H&E explorados neste trabalho.	15
4.1	Valores de acurácias em cada <i>fold</i> de cada execução e as médias correspondentes em porcentagens (%) para os conjuntos UCSB, CR e OED.	18
4.2	Resumos dos desempenhos obtidos a partir das representações CAM.	20
4.3	Médias, desvios-padrão e intervalos de confiança (IC) para avaliação dos índices de relevância nos conjuntos H&E.	22

Sumário

1	Introdução	1
1.1	Justificativas e Motivação	2
1.2	Objetivos	3
1.3	Organização do Trabalho	3
2	Fundamentação Teórica e Trabalhos Relacionados	4
2.1	Redes Neurais Convolucionais	4
2.1.1	Visão Geral sobre Funções de Ativação	5
2.2	Arquitetura <i>Residual Neural Network</i>	5
2.3	Alguns Conceitos sobre Transferência de Aprendizado	7
2.4	Mapas de Ativação de Redes Neurais Convolucionais	8
2.5	Uma Visão Geral sobre Imagens Histológicas	10
2.6	Trabalhos Relacionados	10
2.7	Técnicas para interpretações de modelos CNN: uma visão geral	10
3	Metodologia	12
3.1	Etapa 1: Pré-processamento	13
3.2	Etapa 2: Mapas de Ativação de Classe	13
3.3	Contextos de Aplicação	15
3.4	Métricas para Confirmar os Resultados	16
4	Resultados	18
4.1	Medidas e Mapas de Ativação para os conjuntos H&E	18
5	Conclusão	23

1 Introdução

Algoritmos de aprendizado de máquina tem o potencial de serem utilizados em todas as áreas da medicina, desde a descoberta de medicamentos até a tomada de decisões nas mais diferentes práticas clínicas. Por exemplo, as imagens histológicas constituem uma parte integral de um registro de saúde eletrônico de pacientes, comumente utilizadas para fundamentar diagnósticos e prognósticos médicos. Um diagnóstico tardio ou equivocado de características presentes em imagens pode prejudicar diretamente o prognóstico médico. Neste contexto, o processo de análise de imagens médicas tem sido explorado via sistemas de apoio ao diagnóstico para minimizar os diferentes problemas existentes, além de indicar um diagnóstico correto pelos motivos corretos (KER et al., 2017). Dentre as diversas técnicas de aprendizado de máquina empregadas na análise de imagens histológicas, destacam-se os modelos fundamentados em redes neurais convolucionais (CNN), com aplicações para identificar e prever câncer oral (CAMALAN et al., 2021) e câncer de mama (ARAÚJO et al., 2017), por exemplo. Dentre as limitações que as redes CNN estão submetidas, uma das mais importantes é a necessidade de modelos complementares para tornar processo de decisão compreensível para especialistas.

Neste contexto, a interpretação dos resultados de uma CNN pode ser via uma estratégia de mapas de ativação de classe, do inglês *class activation maps* (CAM). CAM é cada vez mais utilizado para tarefas de *weakly-supervised object localization* (BELHARBI et al., 2022). A vantagem de uma estratégia CAM está em produzir visualizações das regiões mais utilizadas de imagens no processo de classificação, bem como uma pontuação de predição das classes ou índice de relevância (ZHOU et al., 2016). Essas abordagens foram aplicadas com sucesso em diferentes tipos de imagens (HAMMAD; KHOTANLOU, 2022) e (HASSAN et al., 2022).

Apesar dos avanços indicados previamente, ainda é possível investigar o poder discriminativo de uma CNN e se os resultados obtidos por meio quantificações CAM confirmam as classificações, incluindo os destacados pela rede. Por exemplo, é possível conhecer o poder discriminativo de uma camada específica de uma CNN com as indicações de regiões das imagens que embasaram a classificação na camada. Assim, a técnica CAM pode ser útil para conhecer índices de relevância das explicações em imagens histológicas coradas com Hematoxilina-Eosina (H&E). Essas aplicações permitem obter informações relevantes sobre os processos de classificação e reconhecimento de padrões de doenças via imagens H&E. Os conhecimentos obtidos podem embasar e/ou aprimorar o desenvolvimento de sistemas computacionais.

1.1 Justificativas e Motivação

Dentre os diversos tipos de doenças existentes, um dos tipos mais preocupantes na atualidade é o câncer. Exemplos são o câncer oral, chamado também de *oral squamous cell carcinoma* (OSCC), que está entre os dez tipos de câncer mais comuns globalmente, com mais de 500.000 novos casos e 350.000 mortes em 2021 (CAMALAN et al., 2021), câncer de mama e câncer gástrico (responsável por 952.000 mortes em 2019 (CAO et al., 2019)). Por exemplo, no caso do câncer oral, apesar dos diversos tratamentos cirúrgicos e oncológicos, a taxa de sobrevivência de pacientes não melhorou nas últimas quatro décadas. Portanto, a falta de um método clínico objetivo para avaliar lesões orais é uma barreira crítica para a detecção da doença nos estágios iniciais (CAMALAN et al., 2021). Neste contexto, sistemas computacionais de apoio ao diagnóstico podem contribuir com informações relevantes para o estudo de cada tipo de câncer citado previamente, inclusive apoiando os especialistas na prática clínica.

A partir disso, é possível constatar que técnicas computacionais podem contribuir significativamente como uma segunda leitura, apoiando no processo de tomada de decisão e no estudo clínico de cada classe histopatológica. Especificamente neste campo de aplicação, após a coleta do tecido histológico, as amostras são preparadas com o uso de uma etapa de coloração. Nessa etapa, as estruturas presentes nas amostras são essencialmente identificadas em consequência do uso de corantes, que se fixam às estruturas por suas afinidades seletivas a diferentes componentes biológicos. Por fim, a digitalização da lâmina com o tecido permite a obtenção das imagens digitais histológicas ou histopatológicas, comumente utilizadas nos esquemas computacionais (GARTNER, 2020). Logo, o diagnóstico precoce de câncer tem impactos diretos sobre a taxa de sobrevivência de pacientes. Uma das formas de análise e confirmação ocorre por meio da análise microscópica de estruturas celulares e a forma de organização. Entretanto, realizada manualmente, essa tarefa torna-se dispendiosa, complexa e subjetiva, impulsionando o desenvolvimento de técnicas computacionais para enfrentar esse desafio (TALO, 2019) (ELGENDY, 2020) (ZHENG et al., 2018) (GHOSH et al., 2021).

Neste contexto, com o advento da digitalização de amostras histológicas, modelos fundamentados em CNN foram desenvolvidos para auxiliar patologistas, tanto por meio de técnicas de classificação de regiões de interesse, como via abordagens de reconhecimento de padrões em diferentes tipos de câncer e seus graus de severidade (TALO, 2019). Por outro lado, é necessário conhecer como os modelos CNN produzem as soluções, permitindo interpretações mais completas dos resultados e tornando as soluções compreensíveis aos especialistas, em especial no campo de histopatologia. Isso é possível quando o uso de técnicas explicáveis são aplicadas, via mapas CAM,

a fim de compreender os resultados de uma mesma CNN em diferentes de amostras H&E, tais como as aqui exploradas, representativas de displasia oral, câncer colorretal e câncer de mama. Apesar deste tipo de exploração, métricas de desempenhos e índices de relevâncias dos mapas podem ser obtidos para confirmar ou rever os resultados da rede, garantindo uma interpretação quantitativa mais segura para o modelo.

1.2 Objetivos

Neste trabalho é apresentado um estudo envolvendo imagens histológicas H&E via o uso combinado de uma arquitetura de rede neural convolucional com CAM a fim de auxiliar no processo de compreensão e interpretação dos resultados de classificação. Os objetivos específicos foram:

1. Aplicar uma rede neural convolucional com transferência de aprendizado para identificar padrões em imagens histológicas;
2. Utilizar um modelo CAM para gerar mapas representativos de regiões mais exploradas em cada classificação;
3. Extrair métricas e índices de relevância das classificações para confirmar os resultados de uma CNN.
4. Aplicar o modelo em imagens histológicas com relevância científica reconhecida, tais como câncer colorretal, displasia oral e câncer de mama.

1.3 Organização do Trabalho

Esta monografia está dividida em cinco capítulos, incluindo esta introdução. No Capítulo 2 está a fundamentação teórica necessária para o entendimento de algumas técnicas e metodologias computacionais, além de alguns artigos relevantes para embasar a metodologia. No Capítulo 3 estão detalhes metodológicos sobre o uso combinado de uma arquitetura de rede neural convolucional com CAM investigar. No Capítulo 4 estão os resultados obtidos após aplicar o esquema computacional descrito no capítulo anterior. As conclusões foram apresentadas na última seção.

2 Fundamentação Teórica e Trabalhos Relacionados

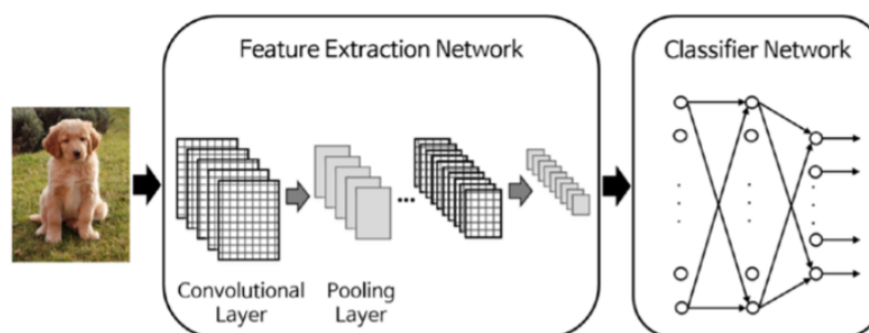
Neste capítulo é apresentada uma visão geral sobre extração de características, redes neurais convolucionais, transferência de aprendizado, processo de validação cruzada, mapas de ativação de classe e imagens histológicas. Adicionalmente, uma seção de trabalhos relacionados também está presente para embasar a investigação envolvendo de mapas de ativação a partir de redes neurais convolucionais.

2.1 Redes Neurais Convolucionais

Redes neurais convolucionais, do inglês *Convolutional Neural Network* (CNN), são um tipo particular de arquitetura de redes neurais para processar dados presentes em múltiplos vetores, comumente explorados para representar imagens coloridas (LECUN et al., 2015). A ideia é ter um computacional capaz de reconhecer eficientemente padrões presentes em imagens. Para realizar esta tarefa, uma CNN é constituída de várias camadas, diferentes e interconectadas, tais como camadas convolucionais, *pooling* e as totalmente conectadas (KULKARNI et al., 2019).

As camadas convolucionais são filtros (matrizes com valores representando pesos) aplicados na imagem de entrada, via um processo de convolução, para produzir os mapas de características, do inglês *feature maps*. Logo, as características obtidas em camada convolucional estão diretamente relacionadas com os pesos que constituem cada filtro, sendo estes definidos em uma etapa de treinamento do modelo CNN (KIM, 2017). A camada de *pooling*, no que lhe concerne, reduz as representações obtidas a partir de cada camada convolucional, permitindo destacar os valores mais relevantes para minimizar a presença *overfitting* e custo computacional (KIM, 2017).

Figura 2.1: Ilustração de uma arquitetura de uma CNN, envolvendo as camadas convolucionais, *pooling* e totalmente conectadas (classificador neural).



Fonte: (KIM, 2017).

2.1.1 Visão Geral sobre Funções de Ativação

Uma CNN é constituída por estruturas para realizar operações matemáticas nos dados de entrada. Para cada neurônio de cada camada, o sinal propagado é o valor de entrada multiplicado pelo peso do neurônio correspondente, somado ao bias do mesmo neurônio. Essas operações expressam relações lineares entre os valores. Para obter representações mais complexas, representativas de relação não-lineares, o sinal propagado deve ser processado a partir de uma função de ativação. Assim, é possível definir condições específicas para representar processos complexos distintos, alterando, por exemplo, a amplitude do sinal propagado (KARLIK; OLGAC, 2011) (SHARMA et al., 2017). Uma CNN sem uma função de ativação resulta em um modelo de regressão linear. Logo, é comum que estas funções sejam aplicadas entre as diferentes camadas de uma CNN. Alguns exemplos de funções de ativação são *rectified linear unit* (ReLU), Sigmoid e SoftMax. A função ReLU (Equação 1) retorna 0 para todos os valores negativos, conservando os demais valores. A função sigmoide (Equação 2) fornece valores no intervalo $[0, 1]$. A função sigmoide foi generalizada para casos não-binários, nomeada SoftMax (Equação 3), útil para tratar problemas multiclasse (SHARMA et al., 2017) e (KARLIK; OLGAC, 2011).

$$ReLU(x) = \max(0, x). \quad (1)$$

$$Sigmoid(x) = \frac{1}{1 + \exp(-x)}. \quad (2)$$

$$Softmax(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^k \exp(x_j)} \text{ para } j = 1, \dots, k. \quad (3)$$

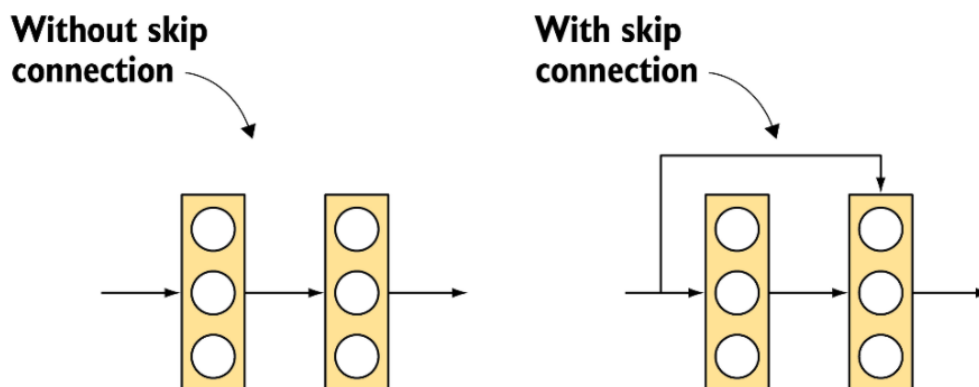
2.2 Arquitetura *Residual Neural Network*

As arquiteturas e modelos CNN evoluíram significativamente ao longo dos anos, seja com pelo aumento do número de camadas ou como estas são organizadas, proporcionando distinções mais precisas (HE et al., 2016) (NANNI et al., 2020). Alguns exemplos são os modelos menos profundos, 19 camadas convolucionais da VGGNet-19 (SIMONYAN; ZISSERMAN, 2014), e os mais profundos, esquema fundamentado na estratégia de *Residual Neural Network* (ResNet), com 50, 101 e 152 camadas. É importante observar que a ResNet com 50 camadas (Resnet-50) é amplamente explorada em esquemas computacionais para apoiar o processo de classificação e reconhecimento de padrões de imagens histopatológicas (NANNI et al.,

2020) (ROBERTO et al., 2021) (TRIPATHI; SINGH, 2020). Isso ocorreu pelo destaque conquistado na competição "*ImageNet Large Scale Visual Recognition Challenge*" (ILSVRC) *ImageNet Large Scale Visual Recognition Challenge*, com um erro de apenas 3,57%, no qual a meta é avaliar algoritmos para detecção de objetos e classificação de imagens no banco de imagens ImageNet (RUSSAKOVSKY et al., 2015).

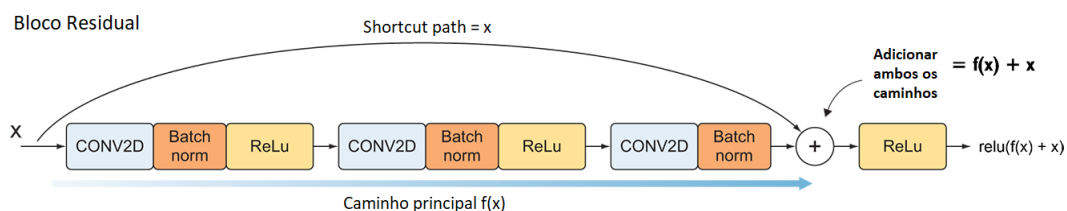
A ResNet-50 venceu a competição ILSVRC 2015 com uma arquitetura explorando a estratégia de blocos residuais (HE et al., 2016), em que uma saída x passa pela conexão *skip connections* (*shortcut path*) e, sem seguida, é adicionada ao caminho principal $f(x)$ (Figura 2.3). Um *shortcut path* é uma estratégia que permite obter conexões por meio de atalhos. Nota-se que $f(x)$ representa a saída da última camada convolucional do bloco, posicionada estrategicamente antes da função de ativação ReLU, resultando em $ReLU(f(x) + x)$. Com isso, a ResNet recorre a conexões residuais para solucionar o problema de degradação durante o treinamento, onde o gradiente é retropropagado para as camadas anteriores. Assim, o modelo aprende uma função de identidade e garante a camada um desempenho similar ou melhor ao da camada anterior (ELGENDY, 2020). Especificamente, a ResNet-50 tem cinco blocos convolucionais empilhados, onde cada bloco residual é repetido ao longo da estrutura. Essa estrutura está ilustrada na Figura 2.4, em que cada cor está associada a um mesmo bloco residual. Na ResNet-50, o primeiro bloco convolucional tem uma única camada convolucional com 64 *kernels* de dimensões 7x7, o segundo (em amarelo) tem três blocos residuais e os próximos são constituídos por quatro, seis e três blocos residuais. Em seguida, existem as camadas *average pooling* e a totalmente conectada (HE et al., 2016).

Figura 2.2: Ilustração da estratégia de *shortcut path* ou *skip connections*.



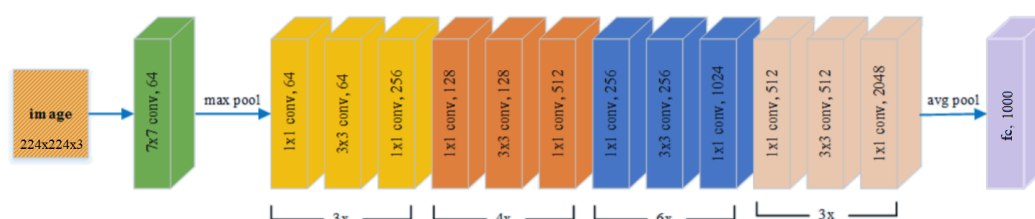
Fonte: (ELGENDY, 2020).

Figura 2.3: Ilustração de um bloco residual na CNN ResNet-50.



Fonte: Adaptado de (ELGENDY, 2020).

Figura 2.4: Ilustração geral da arquitetura ResNet-50.



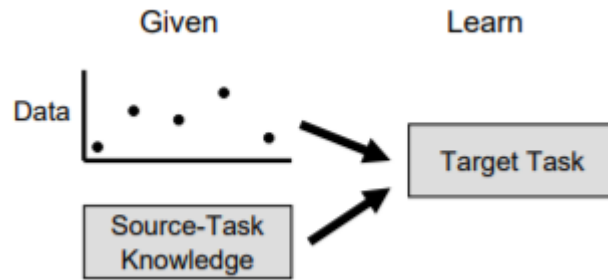
Fonte: Adaptado de (TALO, 2019).

2.3 Alguns Conceitos sobre Transferência de Aprendizado

O uso de modelos fundamentados em *deep learning* exige um volume considerável de dados para treinamento. Esse demanda pode ser contornada com a aplicação da estratégia de aprendizado por transferência, ou *transfer learning*, em que o conhecimento aprendido em um contexto inicial é utilizado para melhorar a aprendizagem em uma tarefa de destino relacionada, mas diferente da inicial, (TORREY; SHAVLIK, 2009).

Modelo do tipo CNN está vinculado ao grupo de algoritmos com aprendizagem indutiva, em que os atributos ou características de entrada são mapeados entre as classes sob avaliação a fim de generalizar os dados e seus relacionamentos. A transferência da aprendizagem indutiva (do mapeamento), de uma tarefa fonte para a classe alvo, pode ser concretizada ajustando o espaço do modelo e corrigindo os valores de bias (variável incluída ao somatório da função de ativação para ajustar a capacidade de aproximação da rede). Isso é possível via substituição da última camada do classificador, refinando os pesos a partir das características generalizadas (ALMARAZ-DAMIAN et al., 2020) e produzindo menor tempo de treinamento, além de minimizar *overfitting* em situações com poucas amostras.

Figura 2.5: Esquema básico de transferência de aprendizado.



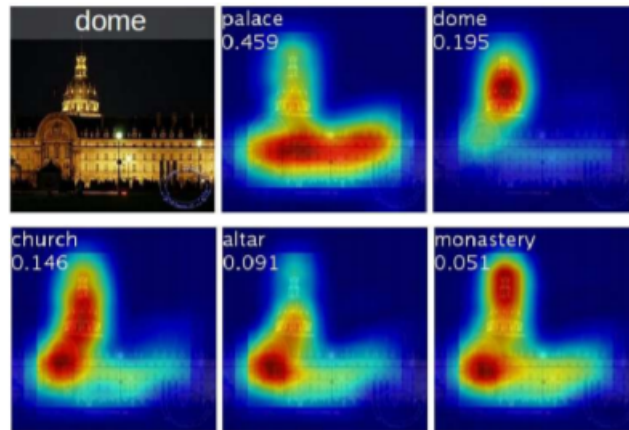
Fonte: (TORREY; SHAVLIK, 2010)

2.4 Mapas de Ativação de Redes Neurais Convolucionais

A extração de características é uma das etapas mais importantes no processo de análise de imagens. A ideia é definir um mapeamento numérico de informações visuais presentes em uma imagem, valores estes comumente organizados como vetores de características. Consequentemente, esse mapeamento é avaliado em um processo de classificação para indicar possíveis padrões (PEDRINI; SCHWARTZ, 2008). Quando as características são definidas via processos internos de uma rede neural convolucional, os resultados são nomeados *deep learned features* e organizados como mapas de ativação de classe (ZHOU et al., 2016).

Um mapa de ativação de classe indica as regiões discriminativas das imagens em processo de análise via uma rede neural convolucional. As representações obtidas relevam as regiões ao projetar os pesos da camada de saída em um arranjo matricial, processo conhecido como mapeamento de classe de ativação (ZHOU et al., 2016). Por exemplo, considerando o resultado presente em um *pooling* médio global, a saída indica a média espacial do mapa de características de cada unidade na última camada convolucional de um modelo CNN. Logo, uma soma ponderada de tais valores é usada para gerar a saída final. Similarmente a este processo, é calculada a soma ponderada dos mapas de características da última camada convolucional para obter o mapa de ativação de classe final (ZHOU et al., 2016). Portanto, em suma, um mapa de ativação de classe ilustra a soma ponderada linear da presença de um determinado padrão visual de uma imagem em diferentes localizações espaciais após as camadas convolucionais em um modelo CNN. O nível de ocorrência de valores é associado a uma cor, produzindo ilustrações similares a mapas de calor (Figura 2.6). O processo descrito está ilustrado na Figura 2.7, em que w_i indicam os pesos de cada saída.

Figura 2.6: Ilustração de um mapa de ativação de classe.



Fonte: (ZHOU et al., 2016).

Figura 2.7: Ilustração do mapeamento de classe de ativação.



Fonte: (ZHOU et al., 2016).

2.5 Uma Visão Geral sobre Imagens Histológicas

A Histologia pode ser definida como o estudo microscópico dos tecidos dos seres vivos, sejam eles vegetais ou animais (GARTNER, 2020). Comumente, este tipo de estudo requer algumas técnicas para o preparo de amostras de tecidos, como a coloração para garantir e destacar propriedades histológicas específicas (GARTNER, 2020). Nesse caso, como muitos dos componentes dos tecidos possuem as mesmas densidades ópticas, eles necessitam ser coloridos para microscópios de luz, majoritariamente com corantes solúveis em água. Os corantes que são mais usados são a hematoxilina, uma base que se prende preferencialmente aos componentes ácidos da célula, como o DNA e o RNA, colorindo-os de azul, e a eosina (H&E), um ácido que se prende aos componentes citoplasmáticos com pH básico, colorindo-os de rosa (GARTNER, 2020). A partir disso, especialistas avaliam as diferentes propriedades e organizações das estruturas por meio de microscópio ótico, permitindo identificar alterações histopatológicas.

2.6 Trabalhos Relacionados

Nesta seção é apresentada uma visão geral sobre trabalhos importantes sobre técnicas utilizadas para proporcionar explicações envolvendo classificações com CNN.

2.7 Técnicas para interpretações de modelos CNN: uma visão geral

Os mapas de ativação de classe do tipo CAM ou sua generalização *Gradient-weighted class activation mapping* (Grad-CAM) são alternativas úteis no campo de *machine learning*, em especial nos modelos que exploram *deep learning*. Por exemplo, (CAMALAN et al., 2021) apresentaram um modelo que explorou redes neurais convolucionais em imagens histológicas representativas de câncer bucal. Os modelos CNN utilizados foram Inception ResNet-v2, ResNet101, Inception-v3 e VGG-16, todos aplicados em dois conjuntos de imagens, um de Piracicaba (Brasil), com 43 amostras, e o outro de Sheffield (Reino Unido), com 30 imagens. Os autores utilizaram uma etapa de aumento geométrico de dados, com rotações e escalas. O melhor resultado foi conquistado via ResNet-101, no banco de imagens de Piracicaba, com acurácias acima de 95%. Os resultados foram interpretados aplicando a técnica de mapas de ativação de classe. De uma maneira geral, mapeamentos de resultados em sistemas *deep learning* contribuíram com aplicações de classificação (MONTALBO, 2022) (WESP et al., 2022) (MAHMOOD et al., 2022). Adicionalmente, *Locally-interpretable model-agnostic explanation* (LIME), outra categoria de mapeamento, também pode gerar explicações com CNN, visto que explora a geração de superpixel, método de segmen-

tação e perturbações locais na imagem de interesse (SALAHUDDIN et al., 2022). Apesar disso, LIME não é uma estratégia que destaca áreas de atenção como CAM, mas um modelo *post hoc* para prover a interpretabilidade de classificadores. Nota-se que estas estratégias foram utilizadas para avaliar imagens médicas, como pode ser observado no campo de imagens histopatológicas (investigação de biomarcadores (SALAHUDDIN et al., 2022) (SEWERYN, 2021) ou sistemas de apoio ao diagnóstico (DABASS et al., 2022) (CHIEN et al., 2022)).

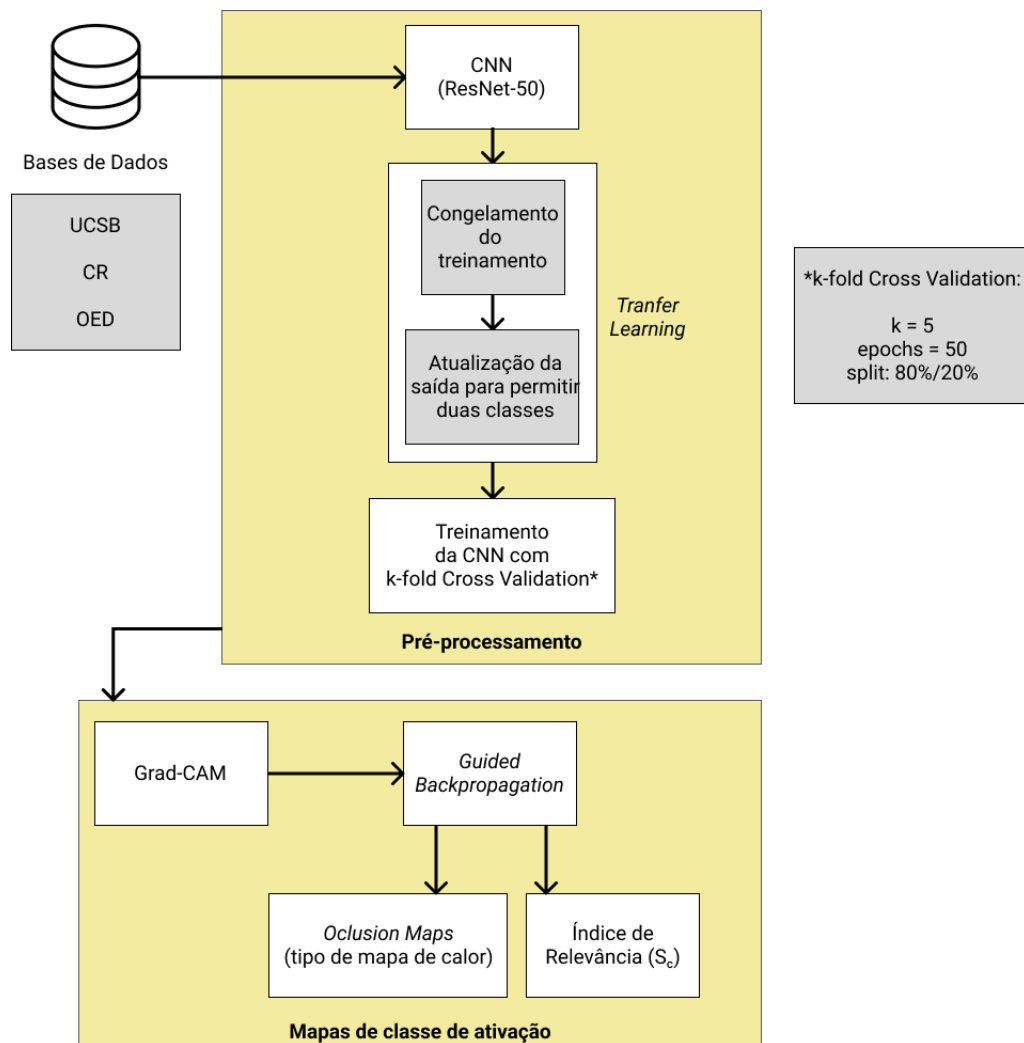
Os mapas proporcionaram interpretações interessantes em outros cenários (CHAO et al., 2022), neste caso, os autores propuseram um método para o reconhecimento de cavitação em bombas de pistão mediante quatro categorias: severa, média, leve e sadia. O método consistiu em realizar o treinamento de redes convolucionais com amostras de espectrogramas para determinar a classificação de severidade da cavidade. Para isso, (CHAO et al., 2022) utilizaram a transformada discreta de Fourier para expandir a representação espacial dos sinais de entrada e aplicaram Grad-CAM para definir representações de máscaras. Cada máscara utilizada em um processo convolucional com a imagem do espectrograma, produzindo uma nova representação do sinal de entrada a fim de melhorar a etapa de treinamento do modelo. Os autores concluíram que as representações Grad-CAM contribuíram com o processo de classificação, permitindo conquistar acurácias de até 89%.

Em complemento aos tipos de aplicações já destacadas, as representações ainda foram investigadas para obter regiões de interesse (DENG et al., 2018) (ZHANG et al., 2022) e, em seguida, combinadas para classificações em cenários diversificados (ABBASNIYA et al., 2022) (YENGEC-TASDEMIR et al., 2022) (NANNI et al., 2020). Observa-se ainda, evoluções importantes para o mapeamento de densidade populacional via imagens (LI et al., 2022), maximização de sinais por meio de eletrocardiogramas (KIM et al., 2022) e classificações no desafio *Fine-grained visual categorization* (KORSCH et al., 2019). Por fim, é possível concluir que representações geradas via mapeamentos contribuem com novas análises no campo de aprendizado de máquina (DABASS et al., 2022) (MONTALBO, 2022), presentes no estudo de imagens médicas (KUMAR et al., 2022) (CHIEN et al., 2022) (YENGEC-TASDEMIR et al., 2022), especialmente para aprimorar o processo de interpretação e validação de resultados (NANNI et al., 2020) (NANNI et al., 2022) (CHAO et al., 2022). No campo da histopatológica, ainda é possível realizar estudos para fornecer informações sobre as relevâncias dos mapas CAM para confirmar os desempenhos dos modelos, em razão de proporcionar destaques de áreas de atenção da CNN, fundamentando novas metodologias ou aprimorando sistemas de apoio ao diagnóstico.

3 Metodologia

O modelo proposto para este trabalho foi definido em duas etapas. A primeira é o pré-processamento da rede neural convolucional, seção 3.1, e a segunda envolve o cálculo e geração dos mapas, seção 3.2. Na Figura 3.1 é apresentada uma ilustração do modelo proposto.

Figura 3.1: Ilustração do modelo proposto contendo as etapas: 1) Pré-processamento; e 2) mapas de classe de ativação.



Fonte: Elaborado pelo autor.

Os algoritmos foram desenvolvidos na plataforma Google Colab em conjunto com Jupyter Notebook. A linguagem de programação utilizada foi o Python, em sua versão 3.9.5. As principais bibliotecas consideradas no trabalho são: Pytorch (versão 1.12.1); Numpy (versão 1.23.2); Matplotlib (versão 3.5.3); e Pandas (versão 1.4.3). A máquina

utilizada para realizar as execuções foi um Notebook MSI com processador Intel Core i7-11800H (2.30GHz), 16 GB de memória RAM (15.7 GB utilizável), GPU Nvidia GeForce RTX 3060 e sistema operacional Windows 10.

3.1 Etapa 1: Pré-processamento

Esta etapa representa os processos iniciais e as preparações da CNN para realizar o modelo de mapas de ativação de classe. A rede neural escolhida foi a ResNet-50, com 50 camadas treináveis (ELGENDY, 2020), em razão de apresentar a menor taxa de erro na competição ILSVRC de 2015 (HE et al., 2016), superando os desempenhos de outras relevantes redes.

É importante destacar que a rede ResNet-50 foi aplicada com os valores de parâmetros obtidos no treinamento realizado na base ImageNet. Os pesos dos parâmetros da camada de classificação foram ajustados a partir da base de imagens de interesse e considerando o método de validação cruzada *K-fold*. Este processo caracterizou o estágio de treinamento fundamento em transferência de aprendizado. Para tanto, o algoritmo identificou a última camada convolucional da rede em questão e realizou o bloqueio do processo de atualização dos pesos (congelamento do treinamento) das demais camadas, deixando somente a última camada disponível para modificações e preservando o aprendizado anterior da CNN. Isto permitiu treinar a rede com as imagens e rótulos de classes do conjunto de imagens em questão. Nos experimentos, foram exploradas envolvendo duas classes *labels*.

O procedimento descrito previamente resultou em um modelo de rede ResNet-50 ajustado para identificar o número de classes (*labels*) sob investigação, por exemplo, benigno e maligno. A generalização desse ajuste foi definida com *k-fold* igual a 5 e um total de 50 *epochs*. O *split* das imagens para a rede foi na proporção de 80% para treinamento e 20% para teste. Estes parâmetros são comumente explorados na Literatura especializada (ROBERTO et al., 2017) (ROBERTO et al., 2021) (CANDELERO et al., 2020) (NANNI et al., 2020). Os resultados na etapa de teste foram coletados na *epoch* que apresentou o menor erro de treinamento. Também, o conjunto de representações CAM e índices de relevância correspondentes foram definidos a partir da *epoch* que produziu o maior valor de acurácia na fase de testes da rede.

3.2 Etapa 2: Mapas de Ativação de Classe

O modelo para geração de mapas de ativação foi realizado com a aplicação do *Gradient-weighted Class Activation Mapping* (Grad-CAM), que utilizou cálculos de gradientes das camadas convolucionais para determinar um mapa de ativação (SELVARAJU et al., 2017). O modelo Grad-CAM é uma generalização do CAM e pode ser

aplicado em diferentes tipos de arquiteturas de CNN e camadas (SELVARAJU et al., 2017). Um ponto importante sobre o Grad-CAM é que este modelo produz diversos tipos de mapas como saída, permitindo visualizações distintas sobre a predição. Um exemplo é o mapa conhecido como *guided backpropagation*, que destaca as características que mais contribuíram para o resultado de classificação. Logo, *occlusion maps* foram obtidos, um mapa de calor (tipo de visualização) das regiões e suas características comumente exploradas para o processo de classificação. Assim, regiões mais observadas receberam as maiores pontuações, identificadas por meio de cores quentes (vermelho, alaranjado e amarelo) (SELVARAJU et al., 2017). Este foi o tipo de mapa explorado neste trabalho. A saída obtida foi uma imagem representativa de um mapa de calor em sobreposição a imagem original, permitindo destacar e identificar as partes de cada imagem que receberam as maiores pontuações.

Os cálculos para esta finalidade foram descritos por (ZHOU et al., 2016). Assim, considerando que $f_k(x,y)$ representa a unidade de ativação k na última camada convolucional com localização espacial (x,y) , foi possível identificar as características visuais da imagem fundamentaram a tomada de decisão. Então, para uma unidade k , o agrupamento médio global foi dado por F^k , via $\sum_{x,y} f_k(x,y)$. Com isso, para uma dada classe c , a entrada para a camada *softmax*, S_c , foi $\sum_k w_k^c F_k(x,y)$, com w_k^c indicando o peso correspondente a classe c para a unidade k . Logo, esta variável permitiu definir a importância de F_k para a classe em questão. Finalmente, a saída da camada *softmax* para a classe c , nomeada como P_c , foi dada por $\frac{\exp(S_c)}{\sum_c \exp(S_c)}$. É importante deixar claro que, para este algoritmo em específico, o bias do softmax foi explicitamente definido como 0, visto que não produz nenhum impacto na classificação (ZHOU et al., 2016).

Ao definir $\sum_{x,y} f_k(x,y)$ na pontuação da classe S_c , temos:

$$S_c = \sum_k w_k^c \sum_{x,y} f_k(x,y) = \sum_{x,y} \sum_k w_k^c f_k(x,y) \quad (4)$$

Em seguida, M_c é considerado o mapa de ativação de classe para a classe c , em que cada elemento espacial deste mapa foi obtido a partir de:

$$M_c(x,y) = \sum_k w_k^c f_k(x,y) \quad (5)$$

Sendo assim, $S_c = \sum_{x,y} M_c(x,y)$ indica diretamente a importância da ativação na grade espacial (x,y) , levando a classificação de uma imagem para a classe c .

Por fim, é necessário ressaltar que o resultado obtido é um mapa de calor para indicar as regiões que mais contribuíram com o processo de classificação, além de valores S_c (índices de relevância): percentuais da imagem representando regiões com e

sem a doença. Logo, por exemplo, para classificações definidas como falsos-positivos, foi possível identificar quantitativamente as regiões da imagem que fundamentaram este resultado por parte da CNN, informação que pode contribuir significativamente com o aprimoramento de sistemas de auxílio ao diagnóstico. O mesmo ocorreu para casos classificados como verdadeiros-positivos e verdadeiros-negativos. Estas são as quantificações fornecidas pelo algoritmo responsável por mapear os resultados via representações CAM.

3.3 Contextos de Aplicação

O modelo foi aplicado em três tipos distintos de imagens histológicas H&E, câncer colorretal (CR), displasia epitelial oral (OED) e mama. Ilustrações de imagens de cada base estão na Figura 3.2.

A base de imagens (CRCRCColorretal) é definida por imagens histológicas derivadas de 16 seções coradas em H&E (*Hematoxylin & Eosin*) do estágio T3 ou T4 de câncer colorretal. As amostras foram rotuladas por histopatologista como benignas ou malignas. A base é constituída por 165 imagens, divididas em 74 imagens de casos benignos e 91 imagens de casos malignos (SIRINUKUNWATTANA et al., 2017).

O conjunto (OEDOEDDisplasia Epitelial Oral) foi produzido via 30 seções histológicas H&E da língua de camundongos submetidos a um carcinógeno (SILVA et al., 2022). Esta base foi aprovada pelo comitê de Ética, protocolo 038/39 da Universidade Federal de Uberlândia, conforme consta em (SILVA et al., 2022). Um total de 148 regiões de interesse foi dividido em duas classes: saudável (74) e displasia severa (74).

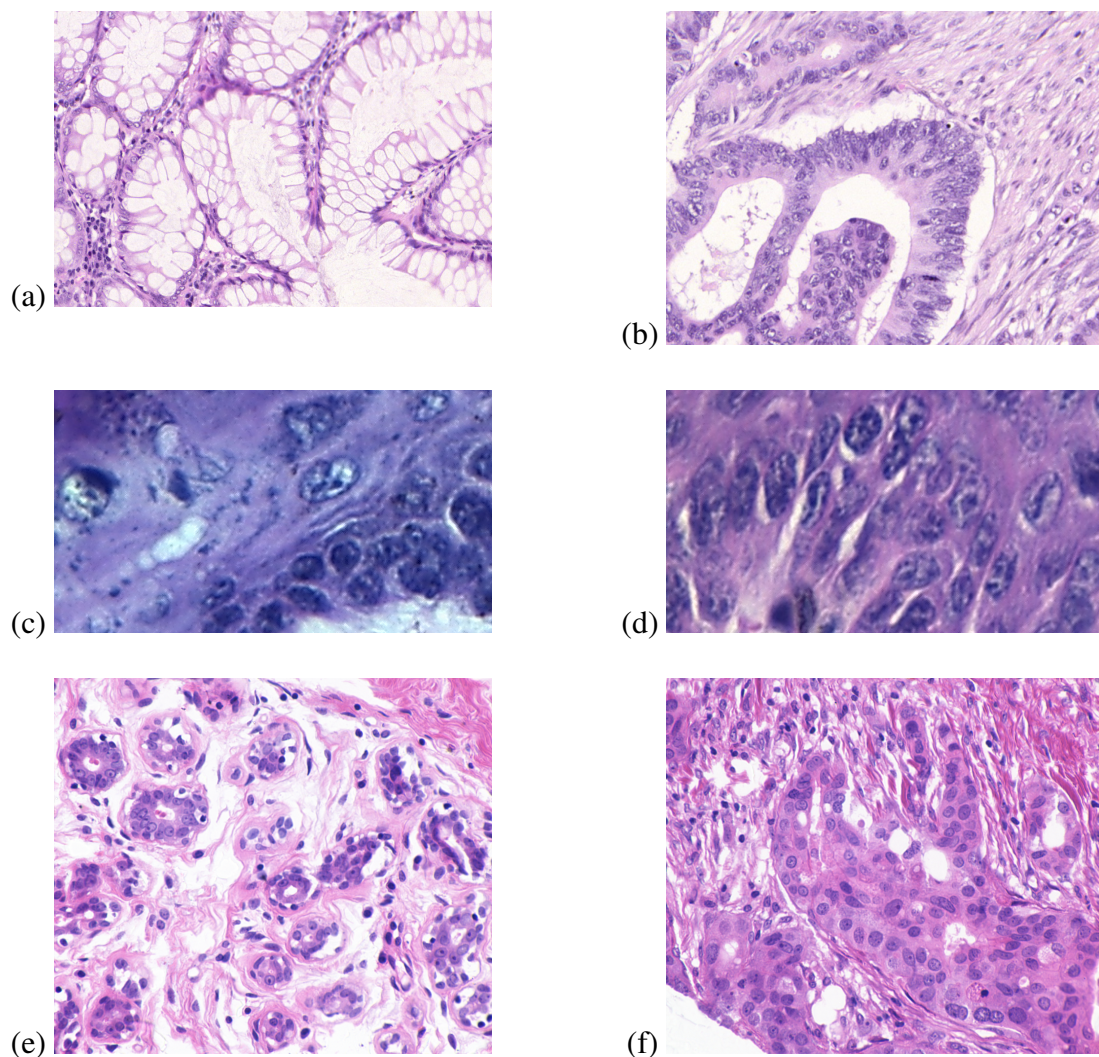
O quarto banco de imagens (UCSB) é composto por 58 imagens de tecido mamário divididos em dois grupos: benigno (32) e maligno (26). A base de imagens é disponibilizada pelo Centro de Informática de Bio-Imagem da Universidade da Califórnia, Santa Barbara (UCSB) (DRELIE GELASCA et al., 2008).

Uma visão geral das bases está na Tabela 3.1, com indicações dos nomes, tipos, quantidade de imagens e classes.

Tabela 3.1: Um resumo dos conjuntos H&E explorados neste trabalho.

Base	Imagem	Quantidade de Imagens	Classes	Resolução
CR	Tumor Colorretal	165	2	567 × 430 a 775 × 522
OED	Displasia Epitelial Oral	148	2	450 x 250
UCSB	Tecido Mamário	58	2	896 × 768

Figura 3.2: Exemplos de imagens H&E: (a) e (b) da base CR; (c) e (d) da base OED; (e) e (f) da base UCSB.



Fonte: Elaborado pelo autor.

3.4 Métricas para Confirmar os Resultados

Os desempenhos da ResNet-50 foram definidos a partir da métrica acurácia, que permite ter uma visão global das classificações (MARTINEZ et al., 2003). A acurácia foi dada pela Equação 6: verdadeiro-positivo (vp), valor positivo classificado corretamente como positivo; verdadeiro-negativo (vn), valor negativo classificado corretamente como negativo; falso-positivo (fp), valor-negativo classificado incorretamente como positivo; e falso-negativo (fn), valor positivo classificado incorretamente como negativo. Essas medidas também foram utilizadas para quantificar as representações CAM. Adicionalmente, para confirmar os desempenhos de distinções da rede, outras medidas foram consideradas: sensibilidade (percentual de resultados positivos den-

tre os resultados positivos), Equação 7; precisão (razão entre vp e a soma de vp com fp), Equação 8; F1-score (média harmônica entre precisão e sensibilidade, permitindo indicar a qualidade do modelo), Equação 9.

$$acuracia = \frac{vp + vn}{vp + fp + vn + fn}. \quad (6)$$

$$sensibilidade = \frac{vp}{vp + fn}. \quad (7)$$

$$precisao = \frac{vp}{vp + fp}. \quad (8)$$

$$F1 - score = 2 \times \frac{precisao \times sensibilidade}{precisao + sensibilidade}. \quad (9)$$

4 Resultados

Neste capítulo, os resultados são apresentados, discutidos e analisados ao aplicar o modelo apresentado no Capítulo 3, com informações sobre os desempenhos obtidos com a rede ResNet-50, visualizações dos mapas para explicar as classificações e valores S_c indicando os percentuais da imagem representando regiões com e sem a doença.

4.1 Medidas e Mapas de Ativação para os conjuntos H&E

O modelo foi aplicado em cada conjunto de imagens com amostras histológicas de câncer colorretal, displasia oral e câncer de mama. Deste modo, as classificações foram realizadas com a rede ResNet-50, usando transferência de aprendizado, e os resultados estão na Tabela 4.1, ilustrando o desempenho em cada *fold* com diferentes composições. O menor desempenho médio foi uma acurácia de 69,40% (base UCSB) e o maior foi de 94,62% (OED, displasia oral). Considerando somente a acurácia como referência, esses resultados indicam um possível potencial do uso combinado da ResNet-50 com transferência de aprendizado, especialmente para classificar e reconhecer padrões nos conjuntos de imagens histológicas representativas de câncer colorretal (acurácia de 92,11%) e displasia oral. Nas condições aqui testadas, a metodologia utilizada não conseguiu conquistar resultados importantes na classificação de padrões presentes na base UCSB. Uma possível explicação é o número reduzido de amostras no conjunto UCSB, com apenas 58 imagens. Por outro lado, em razão do maior número de amostras nas bases CR e OED, é possível inferir que este fato contribuiu para melhorar os ajustes dos pesos das camadas responsáveis pelas classificações. Nestes cenários, os melhores resultados foram obtidos com 50 *epochs*: houve uma estabilização no processo de classificação a partir de 45 *epochs*. Estas condições estão ilustradas na Figura 4.1 para as bases CR (a), OED (b) e UCSB (c).

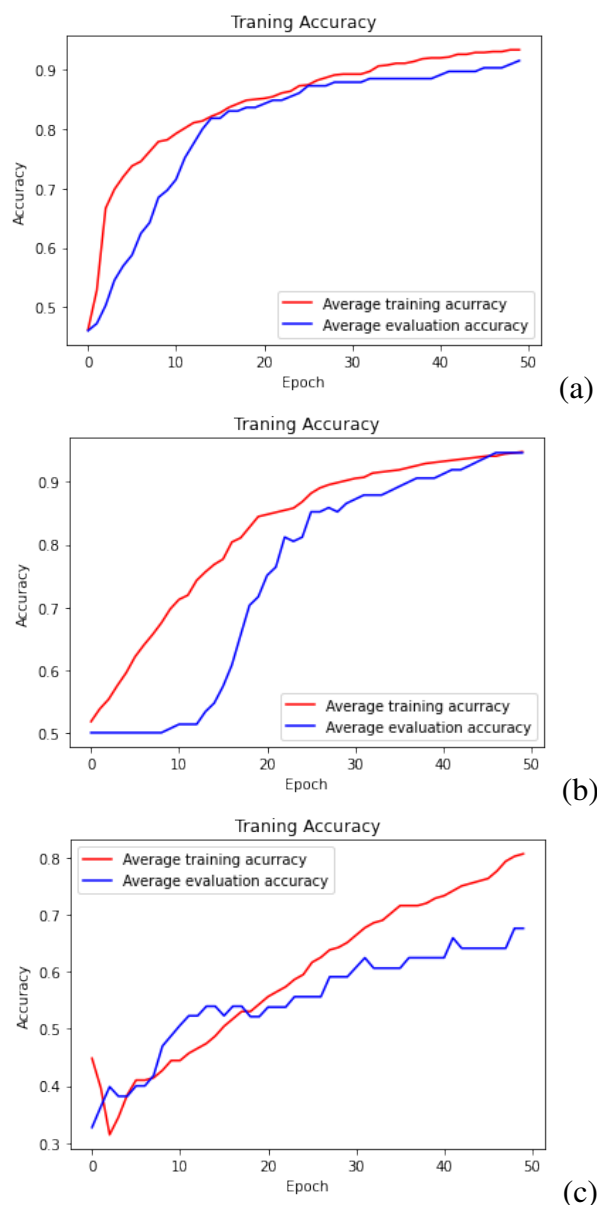
Tabela 4.1: Valores de acurácias em cada *fold* de cada execução e as médias correspondentes em porcentagens (%) para os conjuntos UCSB, CR e OED.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Média
Base UCSB	66,70	41,70	75,00	81,80	81,80	69,40
Base CR	96,97	93,90	90,91	81,8	96,97	92,11
OED	100	90,00	90,00	93,10	100	94,62

Fonte: Elaborado pelo autor.

Para verificar se os valores de acurácias podem ser confirmados a partir das explicações CAM, os valores de precisão, sensibilidade e *F1-score* das representações estão

Figura 4.1: Ilustração da evolução de aprendizado e classificação da ResNet-50 nas imagens H&E: (a) CR; (b) OED; (c) UCSB.



Fonte: Elaborado pelo autor.

na Tabela 4.2, indicando uma visão mais ampla sobre o processo de classificação envolvendo a ResNet-50 e o processo interno de classificação. É importante destacar que essas métricas são calculadas a partir das regiões mais exploradas durante a etapa de quantificação e classificação da CNN.

Observando os resultados disponíveis na Tabela 4.2, é possível notar que a maior precisão (0,93) ocorreu na base CR e a maior sensibilidade no conjunto OED. Apesar disso, observa-se valores discretos de precisão (0,53 para OED e 0,57 para UCSB), inclusive de sensibilidade, 0,48 (UCSB e CR). Para esses dois últimos conjuntos e

Tabela 4.2: Resumos dos desempenhos obtidos a partir das representações CAM.

Métricas	UCSB	CR	OED
precisão	0,57	0,93	0,53
sensibilidade	0,48	0,48	0,98
F1-score	0,52	0,63	0,69

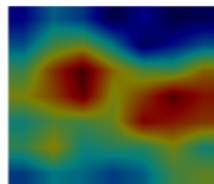
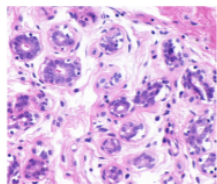
Fonte: Elaborado pelo autor.

considerando o entendimento que a sensibilidade define o quão hábil é o modelo para detectar com sucesso os resultados classificados como positivos, nota-se que as quantificações das representações relevam que não é possível uma confirmação direta da acurácia fornecida pela ResNet-50. Também, quando a precisão é tomada como referência, verifica-se que as explicações não apontam para uma confirmação direta das acurácias conquistadas ara UCSB e OED. Por fim, é importante avaliar os resultados segundo a métrica *F1-score*, média harmônica calculada com base na precisão e na sensibilidade. Neste caso, os resultados das representações foram pouco expressivos, 0,69 na base OED, seguido de 0,63 para CR e somente 0,52 para UCSB. Essas taxas complementam os apontamentos anteriores, com indicativos sobre as limitações das classificações quando confrontadas com as explicações CAM.

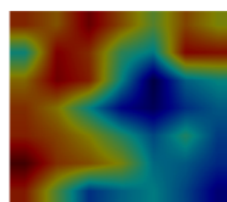
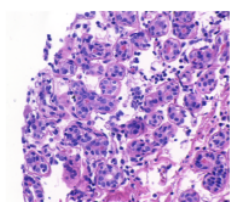
É importante destacar que as explicações foram obtidas diretamente do processo interno de classificação da ResNet-50, via representações CAM, além de índices de relevâncias S_c coletados dos mapas CAM. Logo, cada mapa foi acompanhado por dois valores S_c , quantificando proporções de regiões da imagem pertencentes aos padrões de cada grupo sob investigação. O maior valor S_c é principal explicação da classificação de uma amostra. Exemplos de mapas obtidos para cada conjunto de imagens H&E estão na Figura 4.2, ilustrando representações de cada uma das classes envolvidas no processo de classificação com os índices S_c .

Nos mapas indicados previamente, é possível constatar as principais regiões consideradas para essas predições, com o padrão-ouro e os respectivos índices S_c para explicar os resultados da ResNet-50. As cores mais quentes ilustram as regiões mais exploradas no processo de classificação. Para exemplificar, nas Figuras 6.2(a) e 6.2(c), os mapas com os índices S_c revelam a necessidade de reavaliação por partes dos especialistas, seja para aprimorar o treinamento da CNN ou mesmo o processo de avaliação do especialista. Isso ocorre em razão de divergências nas classificações da ResNet-50 em relação ao padrão-ouro. Mais ainda, essa reavaliação também pode ser aplicada ao exemplo disponível na Figura 6.2(b), visto que os índices são praticamente similares para determinar os padrões de característicos de regiões benigna e maligna. Para uma visão mais amplas sobre os índices S_c , as médias, desvios-padrão e intervalos de con-

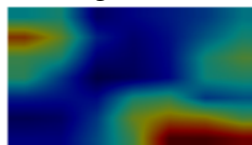
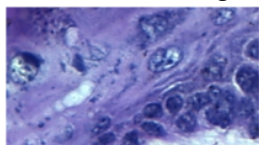
Figura 4.2: Exemplos de imagens histológicas H&E e mapas de ativação correspondentes para explicar as classificações via ResNet-50, envolvendo os conjuntos UCSB ((a) e (b)), OED ((c) e (d)), e CR ((f) e (g)).



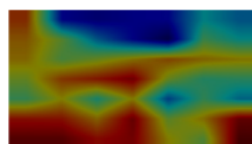
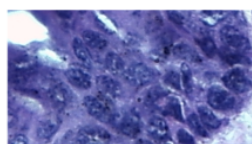
(a) amostra benigna: S_c benigno de 0,44 e maligno de 0,56.



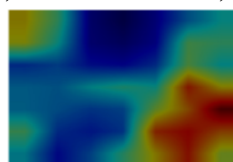
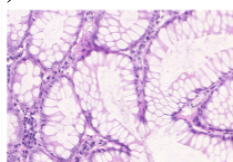
(b) amostra benigna: S_c Benigno de 0,55 e maligno de 0,45.



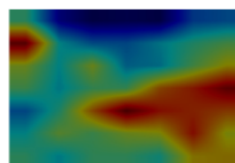
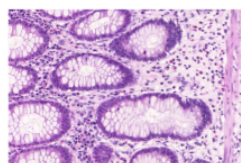
(c) amostra saudável: S_c saudável de 0,36 e severo de 0,64.



(d) amostra saudável: S_c saudável de 0,58 e severo de 0,42.



(e) amostra benigna: S_c benigno de 0,74 e maligno de 0,26.



(f) amostra benigna: S_c benigno de 0,68 e maligno de 0,32.

Fonte: Elaborado pelo autor.

fiança (calculados com um índice de confiança de 95%) são apresentados na Tabela 4.3, considerando cada classe de cada conjunto H&E.

Tabela 4.3: Médias, desvios-padrão e intervalos de confiança (IC) para avaliação dos índices de relevância nos conjuntos H&E.

Índices de relevância	UCSB		CR		OED	
	Benigno	Maligno	Benigno	Maligno	Saudável	Severo
Média	0,52	0,47	0,57	0,42	0,35	0,64
Desvios-padrão	0,07	0,07	0,12	0,12	0,09	0,09
IC (95%)	0,50 a 0,54	0,45 a 0,49	0,55 a 0,59	0,40 a 0,44	0,33 a 0,36	0,64 a 0,65

Fonte: Elaborado pelo autor.

Ao considerar os intervalos S_c presentes na Tabela 4.3, obtidos diretamente das representações CAM, é possível notar possíveis padrões para cada um dos grupos investigados. Esse tipo de informação permite comparações de S_c obtidos de novas classificações ou ainda para fundamentar revisões de pesos da rede, complementando o processo de interpretação do modelo. Análises de representações CAM, como as aqui exploradas, possibilitam uma visão quantitativa das regiões mais relevantes para o modelo, segundo o processo interno de quantificação e classificação da CNN. As médias com os desvios-padrão correspondentes permitem completar o processo de análise, especialmente para confirmar e/ou confrontar as taxas fornecidas pela CNN com as explicações correspondentes. Essas informações são contribuições deste trabalho.

5 Conclusão

O estudo realizado permitiu explorar cenários cientificamente importantes, envolvendo técnicas computacionais para classificar e reconhecer padrões em imagens histológicas H&E representativas de câncer colorretal, displasia oral e câncer de mama.

A partir do modelo aqui desenvolvido e experimentos realizados, foi possível observar que os mapas de ativação podem tornar mais compreensíveis as classificações fornecidas pela ResNet-50, modelo amplamente explorado na Literatura, inclusive para confirmar e/ou confrontar os diferentes resultados obtidos de processos internos do modelo. Nos cenários e condições explorados, as investigações não apontaram para uma confirmação direta das acurácias a partir das métricas obtidas de explicações CAM. Esse tipo de informação em conjunto com um índice de relevância para a classificação contribuem com especialistas e pesquisadores interessados no processo de aprimoramento de sistemas de auxílio ao diagnóstico, possibilitando tornar os resultados mais precisos e compreensíveis para o processo de tomada de decisão. Por exemplo, mesmo com acurácias relevantes para os conjuntos OED (94,62%) e CR (92,11%), os valores $F1 - score$, por exemplo, das representações CAM não ratificaram esses resultados. Mais ainda, as médias e intervalos de confiança dos índices S_c dos grupos investigados permitem um entendimento quantitativo das explicações e do processo de classificação da CNN. Logo, esse é um importante indicativo para reavaliar algumas situações, tais como: amostras com índices S_c sutilmente diferentes; confirmar o enquadramento de amostras indicadas como fp ou fn , via um protocolo de segunda leitura pelos especialistas; rever o processo de treinamento do modelo CNN, buscando minimizar as discrepâncias entre as diferentes métricas adotadas para representar as classificações e explicações CAM; avaliar a necessidade de aumento de dados e os impactos deste procedimento nas explicações CAM. Esses direcionamentos e as informações coletadas são contribuições relevantes deste estudo, visto que foi possível: aplicar com sucesso uma CNN amplamente utilizada na Literatura, combinando transferência de aprendizado, para identificar padrões em imagens histológicas; gerar mapas CAM das regiões mais exploradas durante o processo de classificação da ResNet-50; apresentar métricas e índices de relevância das classificações e explicações para confirmar e/ou confrontar os resultados conquistados.

Referências Bibliográficas

Abbasniya, M. R., Sheikholeslamzadeh, S. A., Nasiri, H., and Emami, S. (2022). Classification of breast tumors based on histopathology images using deep features and ensemble of gradient boosting methods. *Computers and Electrical Engineering*, 103:108382.

Almaraz-Damian, J.-A., Ponomaryov, V., Sadovnychiy, S., and Castillejos-Fernandez, H. (2020). Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures. *Entropy*, 22(4):484.

Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A., and Campilho, A. (2017). Classification of breast cancer histology images using convolutional neural networks. *PloS one*, 12(6):e0177544.

Belharbi, S., Sarraf, A., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). F-cam: Full resolution class activation maps via guided parametric upscaling. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3490–3499.

Camalan, S., Mahmood, H., Binol, H., Araújo, A. L. D., Santos-Silva, A. R., Vargas, P. A., Lopes, M. A., Khurram, S. A., and Gurcan, M. N. (2021). Convolutional neural network-based clinical predictors of oral dysplasia: class activation map analysis of deep learning results. *Cancers*, 13(6):1291.

Candelero, D., Roberto, G. F., do Nascimento, M. Z., Rozendo, G. B., and Neves, L. A. (2020). Selection of cnn, haralick and fractal features based on evolutionary algorithms for classification of histological images. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2709–2716. IEEE.

Cao, G., Song, W., and Zhao, Z. (2019). Gastric cancer diagnosis with mask r-cnn. In *2019 11th International Conference on Intelligent HumanMachine Systems and Cybernetics (IHMSC)*, volume 1, pages 60–63. IEEE.

Chao, Q., Wei, X., Tao, J., Liu, C., and Wang, Y. (2022). Cavitation recognition of axial piston pumps in noisy environment based on grad-cam visualization technique. *CAAI Transactions on Intelligence Technology*.

Chien, J.-C., Lee, J.-D., Hu, C.-S., and Wu, C.-T. (2022). The usefulness of gradient-weighted cam in assisting medical diagnoses. *Applied Sciences*, 12(15):7748.

Dabass, M., Vashisth, S., and Vig, R. (2022). A convolution neural network with multi-level convolutional and attention learning for classification of cancer grades and tissue structures in colon histopathological images. *Computers in Biology and Medicine*, page 105680.

Deng, L., Wang, Y., Han, Z., and Yu, R. (2018). Research on insect pest image detection and recognition based on bio-inspired methods. *Biosystems Engineering*, 169:139–148.

DRELIE GELASCA, E., BYUN, J., OBARA, B., and MANJUNATH, B. S. (2008). Evaluation and benchmark for biological image segmentation. In *2008 15th IEEE International Conference on Image Processing*, pages 1816–1819.

Elgendy, M. (2020). Advanced cnn architectures. <https://livebook.manning.com/book/grokking-deep-learning-for-computervision/chapter-5/v-3/34>. Online; acesso: 30-10-2020.

Gartner, L. P. (2020). *Textbook of histology e-book*. Elsevier Health Sciences.

Ghosh, S., Bandyopadhyay, A., Sahay, S., Ghosh, R., Kundu, I., and Santosh, K. C. (2021). Colorectal histology tumor detection using ensemble deep neural network. *Engineering Applications of Artificial Intelligence*, 100:104202.

Hammad, H. and Khotanlou, H. (2022). Detection and visualization of covid-19 in chest x-ray images using cnn and grad-cam (gccn). In *2022 9th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, pages 1–5. IEEE.

Hassan, H., Ren, Z., Zhou, C., Khan, M. A., Pan, Y., Zhao, J., and Huang, B. (2022). Supervised and weakly supervised deep learning models for covid-19 ct diagnosis: A systematic review. *Computer Methods and Programs in Biomedicine*, page 106731.

He, K., Zhang, X., Ren, S., and Sun, J. (2016a). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

He, K., Zhang, X., Ren, S., and Sun, J. (2016b). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, page 770–778.

Karlik, B. and Olgac, A. V. (2011). Performance analysis of various activation functions in generalized mlp architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, 1(4):111–122.

Ker, J., Wang, L., Rao, J., and Lim, T. (2017). Deep learning applications in medical image analysis. *Ieee Access*, 6:9375–9389.

Kim, J.-K., Jung, S., Park, J., and Han, S. W. (2022). Arrhythmia detection model using modified densenet for comprehensible grad-cam visualization. *Biomedical Signal Processing and Control*, 73:103408.

Kim, P. (2017). Convolutional neural network. In *MATLAB deep learning*, pages 121–147. Springer.

Korsch, D., Bodesheim, P., and Denzler, J. (2019). Classification-specific parts for improving fine-grained visual categorization. In *German Conference on Pattern Recognition*, pages 62–75. Springer.

Kulkarni, U., Meena, S., Gurlahosur, S. V., and Mudengudi, U. (2019). Classification of cultural heritage sites using transfer learning. In *2019 IEEE fifth international conference on multimedia big data (BigMM)*, pages 391–397. IEEE.

Kumar, N., Sharma, M., Singh, V. P., Madan, C., and Mehandia, S. (2022). An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images. *Biomedical Signal Processing and Control*, 75:103596.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

Li, Y.-C., Jia, R.-S., Hu, Y.-X., Han, D.-N., and Sun, H.-M. (2022). Crowd density estimation based on multi scale features fusion network with reverse attention mechanism. *Applied Intelligence*, pages 1–17.

Mahmood, T., Kim, S. G., Koo, J. H., and Park, K. R. (2022). Artificial intelligence-based tissue phenotyping in colorectal cancer histopathology using visual and semantic features aggregation. *Mathematics*, 10(11):1909.

Martinez, E. Z., Louzada-Neto, F., and Pereira, B. d. B. (2003). A curva roc para testes diagnósticos. *Cad. saúde colet.,(Rio J.)*, pages 7–31.

Montalbo, F. J. P. (2022). Diagnosing gastrointestinal diseases from endoscopy images through a multi-fused cnn with auxiliary layers, alpha dropouts, and a fusion residual block. *Biomedical Signal Processing and Control*, 76:103683.

Nanni, L., Ghidoni, S., Brahnam, S., Liu, S., and Zhang, L. (2020a). Ensemble of handcrafted and deep learned features for cervical cell classification. In *Deep Learners and Deep Learner Descriptors for Medical Applications*, pages 117–135. Springer.

Nanni, L., Ghidoni, S., Brahnam, S., Liu, S., and Zhang, L. (2020b). Ensemble of handcrafted and deep learned features for cervical cell classification. In Nanni, L., Brahnam, S., Brattin, R., Ghidoni, S., and Jain, L., editors, *Deep Learners and Deep Learner Descriptors for Medical Applications. Intelligent Systems Reference Library*, volume 186, pages 117–135. Springer.

Nanni, L., Maguolo, G., and Pancino, F. (2020c). Insect pest image detection and recognition based on bio-inspired methods. *Ecological Informatics*, 57:101089.

Nanni, L., Manfè, A., Maguolo, G., Lumini, A., and Brahnam, S. (2022). High performing ensemble of convolutional neural networks for insect pest image detection. *Ecological Informatics*, 67:101515.

Pedrin, H. and Schwartz, W. R. (2008). *Análise de imagens digitais: Princípios, algoritmos e aplicações*, volume 1. Cengage, 1 edition.

Roberto, G. F., Lumini, A., Neves, L. A., and do Nascimento, M. Z. (2021a). Fractal neural network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images. *Expert Systems with Applications*, 166:114103.

Roberto, G. F., Lumini, A., Neves, L. A., and Nascimento, M. Z. (2021b). Fractal neural network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images. *Expert Systems With Applications*, 166:114103.

Roberto, G. F., Neves, L. A., Nascimento, M. Z., Tosta, T. A., Longo, L. C., Martins, A. S., and Faria, P. R. (2017). Features based on the percolation theory for quantification of non-hodgkin lymphomas. *Computers in biology and medicine*, 91:135–147.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., and Ma, S. Huang, Z. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115:211–252.

Salahuddin, Z., Woodruff, H. C., Chatterjee, A., and Lambin, P. (2022). Transparency of deep neural networks for medical image analysis: A review of interpretability methods. *Computers in biology and medicine*, 140:105111.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626.

Seweryn, D. and Osowski, S. (2021). Hierarchical system of gene selection based on deep learning and ensemble approach. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Sharma, S., Sharma, S., and Athaiya, A. (2017). Activation functions in neural networks. *towards data science*, 6(12):310–316.

Silva, A. B., Martins, A. S., Tosta, T. A. A., Neves, L. A., Servato, J. P. S., de Araújo, M. S., DE FARIA, P. R., and DO NASCIMENTO, M. Z. (2022). Computational analysis of histological images from hematoxylin and eosin-stained oral epithelial dysplasia tissue sections. *Expert Systems with Applications*, 193.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*.

Sirinukunwattana, K., Pluim, J. P. W., Chen, H., and X., Q. (2017). Gland segmentation in colon histology images: The glas challenge contest. *MedicalImageAnalysis*, 35:489–502.

Talo, M. (2019). Convolutional neural networks for multi-class histopathology image classification. *ArXiv*, abs/1903.10035.

Torrey, L. and Shavlik, J. (2009). Transfer learning. In Soria, E., Martin, J., Magdalena, R., Martinez, M., and Serrano, A., editors, *Handbook of Research on Machine Learning Applications*.

Torrey, L. and Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global.

Tripathi, S. and Singh, S. H. (2020). Ensembling handcrafted features with deep features: an analytical study for classification of routine colon cancer histopathological nuclei images. *Multimedia Tools and Applications*.

Wesp, P., Grosu, S., Graser, A., Maurus, S., Schulz, C., Knösel, T., Fabritius, M. P., Schachtner, B., Yeh, B. M., Cyran, C. C., et al. (2022). Deep learning in ct colonography: differentiating premalignant from benign colorectal polyps. *European Radiology*, pages 1–11.

Yengce-Tasdemir, S. B., Akay, E., Dogan, S., and Yilmaz, B. (2022). Classification of colorectal polyps from histopathological images using ensemble of convnext variants.

Zhang, H., Zhao, S., Song, Y., Ge, S., Liu, D., Yang, X., and Wu, K. (2022). A deep learning and grad-cam-based approach for accurate identification of the fall armyworm (*spodoptera frugiperda*) in maize fields. *Computers and Electronics in Agriculture*, 202:107440.

Zheng, Y., Yang, C., and Merkulov, A. (2018). Breast cancer screening using convolutional neural network and follow-up digital mammography. In Mahalanobis, A., Ashok, A., Tian, L., and Petrucci, J. C., editors, *Computational Imaging III*, volume 10669, pages 1 – 13. International Society for Optics and Photonics, SPIE.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929.