

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Diagnóstico de Anomalias Estomacais: Utilizando Aprendizado Profundo, Transfer Learning e Machine Learning

Tiago Bernardes Kerr

Monografia - MBA em Inteligência Artificial e Big Data

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Tiago Bernardes Kerr

Diagnóstico de Anomalias Estomacais: Utilizando Aprendizado Profundo, Transfer Learning e Machine Learning

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Fernando Pereira dos Santos

Versão original

São Carlos
2023

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E
PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi, ICMC/USP, com os dados fornecidos pelo(a) autor(a)

S856m	Kerr, Tiago Bernardes Diagnóstico de Anomalias Estomacais: Utilizando Aprendizado Profundo, Transfer Learning e Machine Learning / Tiago Bernardes Kerr ; orientador Fernando Pereira dos Santos. – São Carlos, 2023. 42 p. : il. (algumas color.) ; 30 cm.
	Monografia (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2023.

Tiago Bernardes Kerr

**Diagnóstico de Anomalias Estomacais: Utilizando
Aprendizado Profundo, Transfer Learning e Machine
Learning**

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Original version

**São Carlos
2023**

RESUMO

Kerr, T. B. **Diagnóstico de Anomalias Estomacais: Utilizando Aprendizado Profundo, Transfer Learning e Machine Learning.** 2023. 42p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

O advento da Inteligência Artificial (IA) e, particularmente, do aprendizado profundo (*Deep Learning*), impulsionou avanços notáveis em diversas indústrias. Na área da visão computacional, notadamente na medicina, esses avanços têm revelado eficácia significativa, especialmente no diagnóstico médico por meio de exames de endoscopia. Este trabalho propõe uma solução para o diagnóstico de anomalias estomacais, utilizando imagens de exames endoscópicos. A aplicação de técnicas de Transfer Learning foi empregada, aproveitando três redes neurais convolucionais predefinidas (ResNet50, MobileNet e VGG-19) para extrair características das imagens. Esse processo gerou colunas de características variadas para cada imagem na base de dados. Foram utilizados dois modelos de Machine Learning (KNN e Random Forest) para classificação das anomalias estomacais a partir das características extraídas. Os resultados mostraram que o modelo baseado em Random Forest apresentou melhor desempenho geral, com a MobileNet fornecendo os melhores resultados entre as redes neurais estudadas. A análise da matriz de confusão e métricas específicas, como acurácia, precisão e recall, evidenciou a assertividade e eficácia do modelo Random Forest, especialmente quando utilizado em conjunto com a MobileNet. Os achados deste estudo destacam a promissora viabilidade de utilizar redes neurais convolucionais predefinidas em conjunto com modelos de Machine Learning para diagnóstico de anomalias estomacais. Como próximo passo, a implementação deste modelo em uma aplicação web para classificação automática de imagens e a exploração de outros modelos de Machine Learning representam caminhos interessantes para aprimorar a aplicabilidade e assertividade do modelo proposto.

Palavras-chave: Deep Learning. Transfer Learning. Machine Learning.

ABSTRACT

Kerr, T. B. **One-class Learning.** 2023. 42p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

The emergence of Artificial Intelligence (AI), particularly Deep Learning, has propelled significant advancements across various industries. In the realm of computer vision, notably in medicine, these advancements have demonstrated notable efficacy, especially in medical diagnosis through endoscopy exams. This work proposes a solution for diagnosing stomach anomalies using images from endoscopic examinations. Transfer Learning techniques were employed, leveraging three pre-defined convolutional neural networks (ResNet50, MobileNet, and VGG-19) to extract features from the images. This process generated diverse feature columns for each image in the dataset. Two Machine Learning models (KNN and Random Forest) were utilized for classifying stomach anomalies based on the extracted features. Results indicated that the Random Forest-based model showcased better overall performance, with MobileNet providing the best results among the studied neural networks. Analysis of the confusion matrix and specific metrics such as accuracy, precision, and recall highlighted the accuracy and effectiveness of the Random Forest model, particularly when used in conjunction with MobileNet. The findings of this study underscore the promising feasibility of employing pre-defined convolutional neural networks alongside Machine Learning models for diagnosing stomach anomalies. As a next step, implementing this model in a web application for automatic image classification and exploring other Machine Learning models represent interesting avenues to enhance the applicability and accuracy of the proposed model.

Keywords: Deep Learning. Trasnfer Learning. Machine Leaning.

LISTA DE FIGURAS

Figura 1 – Evolução da IA (KAUL; ENSLIN; GROSS, 2020)	23
Figura 2 – Z-line (POGORELOV <i>et al.</i> , 2017)	25
Figura 3 – Piloro (POGORELOV <i>et al.</i> , 2017)	26
Figura 4 – Ceco (POGORELOV <i>et al.</i> , 2017)	26
Figura 5 – Esofagite (POGORELOV <i>et al.</i> , 2017)	26
Figura 6 – Pólipos (POGORELOV <i>et al.</i> , 2017)	26
Figura 7 – Colite ulcerativa (POGORELOV <i>et al.</i> , 2017)	27
Figura 8 – Pólipos tingidos e levantados (POGORELOV <i>et al.</i> , 2017)	27
Figura 9 – Margens de ressecção tingidas (POGORELOV <i>et al.</i> , 2017)	27
Figura 10 – Resumo arquitetura ResNet (WIKIDOCS, 2023c)	28
Figura 11 – Arquitetura VGG (WIKIDOCS, 2023b)	29
Figura 12 – Arquitetura MobileNet (WIKIDOCS, 2023a)	30
Figura 13 – Matriz de Confusão MobileNet	35
Figura 14 – Matriz de Confusão VGG-19	36
Figura 15 – Matriz de Confusão ResNet50	37

LISTA DE TABELAS

Tabela 1 – Resultados MobileNet	34
Tabela 2 – Resultados VGG-19	35
Tabela 3 – Resultados ResNet50	36
Tabela 4 – Teste qui quadrado entre modelos	37
Tabela 5 – Associação via ANACOR	37

LISTA DE QUADROS

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
abnTeX	ABsurdas Normas para TeX
IBGE	Instituto Brasileiro de Geografia e Estatística
LaTeX	Lamport TeX
USP	Universidade de São Paulo
USPSC	Campus USP de São Carlos

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Contextualização, Motivação	21
1.2	Objetivos	22
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	Trabalhos Relacionados	24
3	METODOLOGIA	25
3.1	Base de dados	25
3.2	Transfer Learning	27
3.2.1	ResNet50	28
3.2.2	VGG-19	29
3.2.3	MobileNet	29
3.3	KNN	30
3.4	Random Forest	30
3.5	Avaliação	31
3.6	Pré Processamento	31
3.7	Comparação de Modelos	32
4	RESULTADOS	33
4.1	MobileNet	34
4.2	VGG19	34
4.3	ResNet50	35
4.4	Comparação de Modelos	36
5	CONCLUSÕES	39
	Referências	41

1 INTRODUÇÃO

O cenário contemporâneo é marcado pelo crescente avanço da Inteligência Artificial (IA) em diversas indústrias, impulsionando otimizações notáveis nos domínios de negócios (LEE, 2019; ARAUJO, 2020). Esse avanço é atribuído, em grande parte, ao advento do *Deep Learning* ou aprendizado profundo, aliado à disponibilidade de significativo poder computacional (ACADEMY, 2022). Uma área de análise de destaque é a visão computacional, revelando eficácia em setores como medicina, agronegócio, marketing digital, entre outros. Aplicações específicas incluem diagnóstico médico, otimização de áreas de plantio, e análise de sentimentos em tempo real.

O impacto positivo dessas soluções é notável, especialmente no campo da visão computacional. Uma implementação efetiva nesse domínio proporciona vantagens substanciais, como a capacidade de avaliar simultaneamente milhares de imagens em tempo reduzido. Em comparação, especialistas humanos demandam consideravelmente mais tempo para realizar a mesma tarefa. Além disso, a homogeneidade na avaliação é assegurada pela visão computacional, minimizando discrepâncias frequentemente observadas entre especialistas humanos. Isso é respaldado por estudos, como os apresentados por (MONGELO *et al.*, 2012), que confirmam a equivalência de assertividade entre modelos de visão computacional e especialistas humanos.

O âmbito da visão computacional aplicada à medicina tem ganhado notoriedade nos últimos anos. Uma pesquisa no Google Scholar com o termo "*computer vision for medicine*" resulta em aproximadamente 4.380.000 resultados (considerando todo o período). Desde 2019, foram registrados cerca de 37.000 resultados. Temáticas em evidência englobam abordagens gerais que aplicam aprendizado profundo à medicina, bem como aquelas relacionadas à COVID-19.

O presente trabalho propõe uma solução para o diagnóstico de anomalias estomacais, utilizando exames de endoscopia. O processo compreende a extração de características por meio de três redes neurais distintas, seguido pela aplicação de dois algoritmos de Machine Learning (KNN e Random Forest). A comparação dos resultados visa identificar a abordagem mais eficaz para o propósito proposto.

1.1 Contextualização, Motivação

A crescente aplicação da Inteligência Artificial (IA) em exames de imagens médicas representa uma tendência ascendente. Trabalhos específicos, como o presente, que se concentram em um tipo particular de exame, contribuem significativamente para a expansão do uso da IA na área médica. A falta de trabalhos que utilizam classificação de imagens a

partir de imagens de endoscopia utilizando transfer learning é um ganho para identificar se essa abordagem melhora de forma significativa os resultados. Essa abordagem não apenas impulsiona a eficiência, mas também promove a padronização dos serviços, resultando em benefícios tangíveis para a população em geral, dada a abrangência do impacto da saúde.

1.2 Objetivos

A questão de pesquisa centraliza-se na criação de uma solução para a classificação de 8 tipos de condições/doenças estomacais separados em três grupos: Marcos Anatômicos (Z-line, Piloro e Ceco), Achados Patológicos (Esofagite, Pólipos e Colite ulcerativa) e Remoção de Pólipos (Pólipos tingidos e levantados 6, Margens de ressecção tingidas), a partir de imagens obtidas em exames endoscópicos, utilizando algoritmos de *Deep Learning* e *Machine Learning*. O objetivo primordial é desenvolver uma abordagem que aprimore a precisão e eficácia no diagnóstico dessas doenças, contribuindo assim para o avanço da aplicação da IA na medicina.

2 FUNDAMENTAÇÃO TEÓRICA

Visão computacional, conforme definido por (MILANO; HONORATO, 2010), refere-se à capacidade de uma máquina interpretar imagens, ou seja, como um computador percebe seu ambiente, analisando dados significativos extraídos de imagens provenientes de diversas fontes. Esse processo possibilita a manipulação e categorização de objetos presentes em uma imagem (BALLARD, 1982).

Segundo (MILANO; HONORATO, 2010), a velocidade com que o olho humano interpreta informações visuais é notavelmente rápida, atribuída à complexidade do córtex visual no cérebro. Esse aspecto do processamento cerebral tornou-se um foco de estudos visando entender e replicar seu funcionamento na visão computacional.

O aprendizado na visão computacional, inicialmente direcionado pela apresentação de imagens a algoritmos para a classificação de novas imagens com base nos pixels, tornou-se mais eficaz com o advento do *Deep Learning*. Esse avanço ocorreu após décadas desde o início da aplicação de Inteligência Artificial na medicina, datando de 1950 com o *Turing Test*, como ilustrado na Figura 1 (KAUL; ENSLIN; GROSS, 2020).

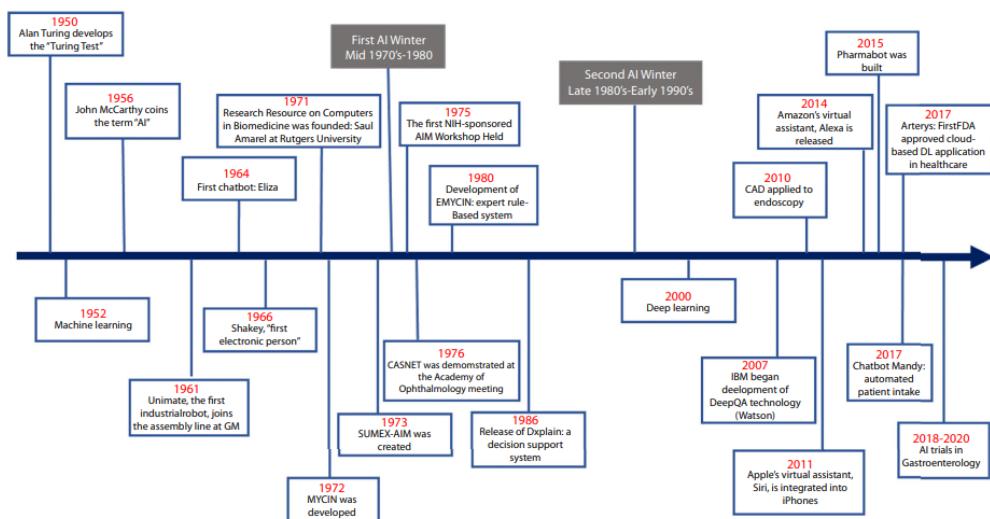


Figura 1 – Evolução da IA (KAUL; ENSLIN; GROSS, 2020)

(KAUL; ENSLIN; GROSS, 2020) destacam outros fatores cruciais, como o avanço do processamento computacional, especialmente com o uso de Unidades de Processamento Gráfico (GPU), e a disponibilidade de grandes conjuntos de dados abertos e categorizados. Esses elementos propiciaram uma explosão de estudos na interseção entre Inteligência Artificial e medicina. Um exemplo notável é o trabalho de (HU *et al.*, 2018), que abrange diversas áreas médicas, utilizando algoritmos como *convolutional neural networks*, *fully convolutional networks*, *auto-encoders*, and *deep belief networks* para diagnosticar vários

tipos de câncer. O estudo oferece sugestões para melhorias futuras e evidencia resultados positivos na detecção de diferentes tipos de câncer.

2.1 Trabalhos Relacionados

A crescente aplicação de técnicas de visão computacional na detecção de anomalias médicas tem demonstrado um notável avanço. Este estudo busca aprimorar os modelos para a classificação de diagnósticos, integrando metodologias estatísticas na comparação de diferentes algoritmos.

A base de dados Kvasir (POGORELOV *et al.*, 2017), de acesso público, tem sido amplamente utilizada em diversos estudos. Uma busca por "Kvasir image classification" no Google Scholar retornou 1550 resultados até o momento, abrangendo pesquisas entre 2019 e 2023. No entanto, a maioria desses trabalhos não está disponível em português. Referências como (LIMA *et al.*, 2023) e (THOMAZ, 2020) exploraram essa mesma base de dados, porém adotaram abordagens e técnicas distintas das propostas neste estudo. O avanço rápido da Inteligência Artificial motiva a busca por resultados mais eficazes ou com melhor desempenho em comparação com estudos anteriores que utilizaram essa base de dados.

Em 2015, cerca de 8 milhões de mortes foram relacionadas a doenças gastrointestinais (CHAN *et al.*, 2019). Diversas pesquisas abordam as classes estudadas neste trabalho de maneira individual. Por exemplo, (JR, 2006) examina o ceco, identificando-o como causador de vários sintomas, como dor abdominal intermitente, distensão, empanzinamento, cólicas, entre outros. Já (MAGALHÃES, 2019) analisa os pólipos, que são detectados em aproximadamente 1% a 4% dos exames e demonstra que, embora a maioria seja benigna e não necessite de intervenção, alguns apresentam potencial maligno. Há outros estudos que abrangem as doenças apresentadas neste trabalho.

Essas doenças são comuns e exigem tempo e atenção por parte do examinador para avaliação da condição, levando em média 7 dias úteis para obtenção dos resultados dos exames. Soluções altamente precisas e rápidas seriam extremamente benéficas para auxiliar os profissionais da saúde, possibilitando avaliações mais céleres e permitindo que os pacientes iniciem o tratamento mais prontamente.

3 METODOLOGIA

3.1 Base de dados

Os dados analisados correspondem à base de dados pública denominada Kvasir (POGORELOV *et al.*, 2017), um conjunto de imagens classificado para auxiliar na detecção computadorizada de doenças gastrointestinais. Essas imagens foram capturadas por um dispositivo endoscópico pertencente ao Vestre Viken Heath Trust (VV), localizado na Noruega, uma entidade composta por quatro hospitais que atendem a uma população de 470.000 pessoas.

A base de dados é composta por 8.000 imagens, categorizadas em 8 classes distintas, com 1.000 imagens por classe. As classes são agrupadas em três categorias principais: Marcos Anatômicos (que incluem Linha Z, Piloro e Ceco), Achados Patológicos (como Esofagite, Pólipos e Colite Ulcerativa) e Remoção de Pólipos (englobando Pólipos Tingidos e Levantados, e Margens de Ressecção Tingidas). A resolução das imagens na base de dados varia de 720x576 a 1920x1072 pixels. As Figuras 2 a 9 ilustram amostras representativas de cada classe incluídas neste estudo.

Marcos Anatômicos: É um atributo visível no trato gastrointestinal que é facilmente identificável através do endoscópio.

A linha Z é definido como o local de transição entre o esôfago e o estômago. Um exemplo de sua imagem é mostrada na Figura 2



Figura 2 – Z-line (POGORELOV *et al.*, 2017)

Piloro é definido como a área ao redor da abertura do estômago até a primeira parte do intestino delgado (duodeno). Um exemplo de sua imagem é mostrada na Figura 3



Figura 3 – Piloro (POGORELOV *et al.*, 2017)

Ceco é a parte mais proximal do intestino grosso. Um exemplo de sua imagem é mostrada na Figura 4



Figura 4 – Ceco (POGORELOV *et al.*, 2017)

Achados Patológicos: É um atributo anormal no trato gastrointestinal

Esofagite é uma inflamação do esôfago visível como uma ruptura na mucosa esofágica em relação à linha Z. Um exemplo de sua imagem é mostrada na Figura 5



Figura 5 – Esofagite (POGORELOV *et al.*, 2017)

Pólipos são lesões no intestino detectáveis à medida que a mucosa cresce. Um exemplo de sua imagem é mostrada na Figura 6



Figura 6 – Pólipos (POGORELOV *et al.*, 2017)

Colite ulcerativa é uma doença inflamatória crônica que afeta o intestino grosso. Um exemplo de sua imagem é mostrada na Figura 7



Figura 7 – Colite ulcerativa (POGORELOV *et al.*, 2017)

Remoção de pólipos: Os pólipos no intestino grosso são associados como precursores de câncer, por isso sua remoção é necessária.

A Figura 8 ilustra um exemplo de pólipos tingidos e levantados por injeção de solução salina e índigo carmim.

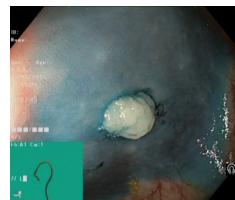


Figura 8 – Pólipos tingidos e levantados (POGORELOV *et al.*, 2017)

Margens de ressecção tingidas são importantes para avaliar se o pólipos foi totalmente removido ou não. O tecido residual do pólipos pode levar ao crescimento contínuo e, no pior dos casos, ao desenvolvimento de malignidade. Um exemplo de sua imagem é mostrada na Figura 9



Figura 9 – Margens de ressecção tingidas (POGORELOV *et al.*, 2017)

3.2 Transfer Learning

O Transfer Learning é um paradigma no campo de aprendizado de máquina reconhecido por sua habilidade em aprimorar o desempenho de modelos em tarefas específicas, alavancando informações obtidas em tarefas correlatas anteriores. Esta abordagem visa transferir o conhecimento adquirido em uma 'Tarefa 01' para melhorar a capacidade de generalização em uma 'Tarefa 02'. Essencialmente, o Transfer Learning explora os pesos

aprendidos por uma rede neural em uma tarefa anterior para melhorar a execução de uma nova tarefa.

A principal motivação para aplicar essa técnica reside na capacidade de utilizar o conhecimento obtido a partir do treinamento de uma rede neural com um extenso conjunto de dados e aplicá-lo em uma nova tarefa com um conjunto de dados limitado, evitando assim iniciar o processo de aprendizado desde o início. Essa estratégia é de particular importância em situações em que os recursos computacionais são restritos, especialmente ao lidar com problemas de treinamento em conjuntos de dados volumosos.

No contexto do Transfer Learning, diferentes abordagens podem ser implementadas. Neste estudo, optou-se pela estratégia de Extração de Características (Feature Extraction), onde as representações de alto nível aprendidas por três redes neurais distintas (ResNet50, MobileNet e VGG-19) na penúltima camada de saída foram utilizadas como entrada em um novo modelo de machine learning. Em vez de reajustar completamente o modelo pré-treinado, as características extraídas são empregadas para treinar um novo modelo específico para a tarefa em questão.

3.2.1 ResNet50

As Redes Residuais (ResNets) representam uma arquitetura avançada de redes neurais profundas utilizada em aprendizado profundo (deep learning). Introduzidas por (HE *et al.*, 2016), as ResNets foram projetadas para mitigar o desafio comum de treinar redes profundas, conhecido como o problema de desvanecimento do gradiente. Essa abordagem inovadora incorpora conexões de atalho, permitindo que a informação flua diretamente através da rede sem ser afetada negativamente pelos efeitos adversos da retropropagação em camadas mais profundas. A utilização de blocos residuais, nos quais a entrada original é somada à saída de uma camada intermediária, facilita a aprendizagem de representações mais eficazes e contribui para o treinamento mais estável e eficiente de redes neurais extremamente profundas. As ResNets tornaram-se um marco importante em visão computacional e outras tarefas de aprendizado profundo, demonstrando melhorias significativas na precisão e eficiência do treinamento em comparação com arquiteturas convencionais. O resumo de sua arquitetura é mostrada em Figura 10

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112			7×7, 64, stride 2		
				3×3 max pool, stride 2		
conv2.x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3.x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4.x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5.x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1			average pool, 1000-d fc, softmax		
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figura 10 – Resumo arquitetura ResNet (WIKIDOCs, 2023c)

3.2.2 VGG-19

A VGG-19 é uma arquitetura de rede neural convolucional desenvolvida pela Visual Geometry Group (VGG) do Departamento de Engenharia Eletrônica e Elétrica da Universidade de Oxford. Proposta por (SIMONYAN; ZISSERMAN, 2014), a VGG-19 é caracterizada por sua profundidade notável, consistindo em 19 camadas convolucionais e totalmente conectadas. A rede é conhecida por sua simplicidade conceitual, baseada na repetição de blocos de camadas convolucionais 3x3 e camadas de pooling, mantendo um número relativamente constante de filtros em cada camada. Essa uniformidade facilita a interpretação e análise da rede. A VGG-19 alcançou sucesso notável em competições de reconhecimento de imagens, destacando-se pela capacidade de extrair características complexas e hierárquicas das imagens, contribuindo para seu papel proeminente no avanço da pesquisa em visão computacional. Apesar de sua profundidade, a VGG-19 tornou-se uma referência e uma base para arquiteturas mais avançadas, influenciando significativamente o desenvolvimento subsequente de redes neurais convolucionais. Os parâmetros de 5 tipos de ResNet, incluindo a usada neste trabalho ResNet50 são vistos em Figura 11

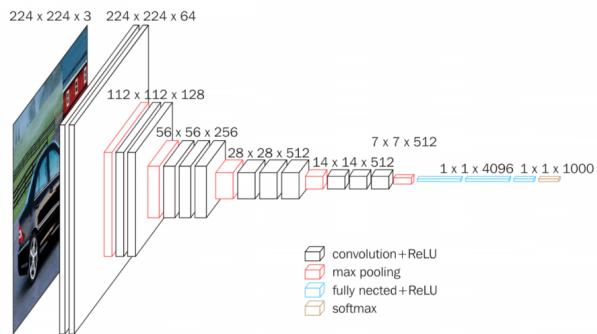


Figura 11 – Arquitetura VGG (WIKIDOCS, 2023b)

3.2.3 MobileNet

A MobileNet é uma arquitetura de rede neural convolucional projetada especificamente para otimizar a eficiência computacional e o desempenho em dispositivos com recursos limitados, como smartphones e dispositivos embarcados. Desenvolvida por (HOWARD *et al.*, 2017), a MobileNet utiliza um conjunto de técnicas inovadoras, como convoluções separáveis em profundidade e convoluções lineares, visando reduzir drasticamente o número de parâmetros e operações computacionais sem comprometer significativamente a precisão da rede. Essa abordagem permite que a MobileNet atinja uma eficiência computacional notável, tornando-a especialmente adequada para implantações em dispositivos móveis com restrições de recursos, onde a eficiência energética e o espaço de armazenamento são considerações críticas. A arquitetura MobileNet contribuiu significativamente para avanços em visão computacional em dispositivos móveis e aplicações de Internet das Coisas (IoT), oferecendo uma solução viável para aplicações práticas que demandam inferência de redes

neurais profundos em ambientes com recursos limitados. Sua arquitetura é vista em Figura 12

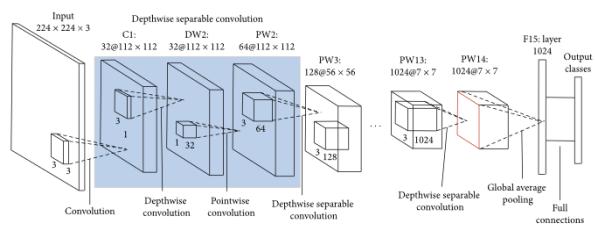


Figura 12 – Arquitetura MobileNet (WIKIDOCs, 2023a)

3.3 KNN

O k-Nearest Neighbors (KNN) (HART; STORK; DUDA, 2000) é um algoritmo de classificação e regressão no campo de aprendizado de máquina supervisionado. Essa abordagem é fundamentada no princípio da similaridade, onde a classificação ou predição de um novo ponto de dados é determinada pela maioria das classes ou valores dos k pontos de dados mais próximos no espaço de características. A proximidade é geralmente medida usando métricas de distância, como a distância euclidiana. O parâmetro k , que define o número de vizinhos considerados, é uma variável crucial na configuração do algoritmo, impactando a sensibilidade do modelo à variação local nos dados. O KNN é não paramétrico e adaptativo, pois não faz suposições explícitas sobre a distribuição subjacente dos dados. Embora seja simples e intuitivo, o KNN pode ser computacionalmente caro para conjuntos de dados grandes e é sensível a atributos irrelevantes ou ruidosos. No entanto, sua aplicabilidade em uma variedade de contextos e a ausência de suposições sobre a natureza dos dados tornam o KNN uma ferramenta versátil e amplamente utilizada em várias áreas de reconhecimento de padrões e análise de dados.

3.4 Random Forest

A Floresta Aleatória (Random Forest) é um algoritmo de aprendizado de máquina que se baseia em um conjunto de árvores de decisão para realizar tarefas de classificação ou regressão. Proposta por (BREIMAN, 2001), a Random Forest introduz aleatoriedade durante o treinamento, utilizando subamostras aleatórias do conjunto de dados e selecionando características de forma aleatória em cada nó de decisão de cada árvore. Isso promove uma diversidade entre as árvores individuais, evitando sobreajuste e melhorando a generalização para novos dados. Cada árvore na floresta opera de maneira independente durante a inferência, e as previsões são combinadas para formar uma decisão final robusta. Por outro lado, uma Árvore Aleatória (Random Tree) refere-se a uma única árvore de decisão treinada com amostras de dados aleatórias, mas sem a adição de aleatoriedade

nas características. Embora as árvores aleatórias também possam ser eficazes, a força distintiva da Random Forest reside na combinação de múltiplas árvores, proporcionando maior estabilidade e desempenho geralmente superior em comparação com uma única árvore aleatória.

3.5 Avaliação

A matriz de confusão é uma tabela que descreve o desempenho de um modelo de classificação, comparando as previsões do modelo com os rótulos verdadeiros dos dados. Ela é composta por quatro elementos: verdadeiros positivos (TP), falsos positivos (FP), verdadeiros negativos (TN) e falsos negativos (FN). A acurácia (accuracy) é a proporção de previsões corretas em relação ao total de previsões e é calculada como $\frac{(TP+TN)}{(TP+FP+TN+FN)}$. A precisão (precision) é a proporção de verdadeiros positivos em relação à soma de verdadeiros positivos e falsos positivos, ou seja, $\frac{TP}{(TP+FP)}$. O recall, também conhecido como sensibilidade ou taxa de verdadeiros positivos, é a proporção de verdadeiros positivos em relação à soma de verdadeiros positivos e falsos negativos, calculada como $\frac{TP}{(TP+FN)}$. O F1-score é uma métrica que combina precisão e recall, representando a média harmônica entre essas duas medidas e sendo calculado como $2 * \frac{(precision \times recall)}{(precision + recall)}$. Cada uma dessas métricas fornece uma perspectiva única sobre o desempenho do modelo em diferentes aspectos, sendo úteis para avaliar sua eficácia em tarefas de classificação.

3.6 Pré Processamento

O resultado do Transfer Learning, no qual as redes neurais convolucionais ResNet50, VGG-19 e MobileNet foram aplicadas para extrair características das imagens, gerou para a ResNet50 38.400 colunas de características, para a VGG-19 produziu 30.000 colunas e a MobileNet 50.176 colunas de características para cada imagem na base de dados e esse resultado serviu como entrada para os modelos de machine learning (KNN e Random Forest).

No processo de pré-processamento, os dados foram divididos em conjuntos de treinamento e teste com uma proporção de 80% para treinamento e 20% para teste. Para garantir uma distribuição uniforme das classes nos dados de treinamento e validação durante o treinamento dos modelos, foi empregada a técnica de validação cruzada estratificada (ATKESON; MOORE; SCHAAAL, 1997). Isso foi realizado com o intuito de mitigar qualquer viés resultante de desequilíbrios nas classes.

Os modelos utilizados, Random Forest e K-Nearest Neighbors, foram configurados com parâmetros específicos: o Random Forest foi configurado com 100 estimadores, enquanto o KNN foi ajustado para considerar 5 vizinhos mais próximos. Em seguida, o conjunto de treinamento foi submetido a um processo de validação cruzada para avaliar a capacidade de generalização do modelo KNN. Posteriormente, ambos os modelos foram

empregados para fazer previsões nos dados de teste, gerando assim resultados preditivos para a avaliação de suas performances.

3.7 Comparação de Modelos

O teste qui-quadrado é uma ferramenta estatística utilizada para avaliar a independência entre variáveis categóricas. Ele se baseia na comparação entre as frequências observadas e as frequências esperadas, permitindo determinar se há uma relação significativa entre as variáveis analisadas. As hipóteses nula (H_0) e alternativa (H_1) neste contexto envolvem a independência das variáveis e a presença de associação entre elas, respectivamente (SIEGEL, 1981).

No contexto de comparação de três modelos através de suas matrizes de confusão, o teste qui-quadrado pode ser empregado para analisar se a assertividade na classificação das classes é significativamente diferente entre esses modelos. Ao somar os acertos e erros das previsões de cada modelo a partir das matrizes de confusão, é possível obter as frequências observadas para cada categoria. Em seguida, ao estimar as frequências esperadas sob a hipótese de independência entre os modelos e as classes, o teste qui-quadrado pode ser aplicado para determinar se existe uma associação significativa (no presente trabalho adotou que p-valores menores do que 0,05 são considerados significativos) entre o modelo utilizado e a assertividade na classificação das classes.

Análise de Correspondência (Anacor) é uma técnica estatística multivariada utilizada para explorar e visualizar associações entre variáveis categóricas em um conjunto de dados. Quando o p-valor associado ao teste qui-quadrado realizado na Anacor é significativo, isso indica que existe uma associação estatisticamente significativa entre as categorias das variáveis estudadas. Em outras palavras, rejeitar a hipótese nula (H_0) no teste qui-quadrado na Anacor sugere que há uma relação não aleatória entre as variáveis.

A significância estatística do p-valor obtido neste teste indica que existe um padrão de associação entre as categorias das variáveis envolvidas na análise. Isso significa que a disposição das categorias em diferentes dimensões nos gráficos de Anacor não é aleatória, mas sim influenciada por uma relação subjacente entre elas. A interpretação dos resultados significativos na Anacor com um p-valor baixo indica a presença de associações relevantes e não aleatórias entre as categorias das variáveis estudadas. A interpretação na Análise de Correspondência (Anacor) ocorre da seguinte maneira: o residual padronizado do teste qui-quadrado representa o grau de associação. Para um nível de confiança de 95%, valores do residual padronizado acima de 1,96 indicam associações entre as categorias das variáveis analisadas.

4 RESULTADOS

O processo de Feature Extraction tem um resultado para cada rede neural, tanto em valores, quanto na quantidade de variáveis e no tempo de processamento, onde Resnet50 gerou 38400 variáveis levando de menos de 1 a 8 segundos por imagem, VGG19 25089 variáveis levando de menos de 1 a 5 segundos por imagem e MobileNet 50176 variáveis levando até 1 segundo por imagem. Para cada saída das redes neurais foram aplicados dois algoritmos, KNN e Random Forest. Os resultados de cada rede são mostrados abaixo, na matriz de confusão os códigos são em relação às seguintes classes: **Z-line 0, Piloro 1, Ceco 2, Esofagite 3, Pólipos 4, Colite ulcerativa 5, Pólipos tingidos e levantados 6 e Margens de ressecção tingidas 7**. Também foi testado um modelo de redução de dimensionalidade antes da aplicação dos modelos de Machine Learning, mas os resultados não ficaram interessantes, com as previsões ficando com métricas baixas em todas as classes e não são apresentados no presente trabalho.

4.1 MobileNet

É possível perceber pela matriz de confusão ilustrada na Figura 13, é possível perceber que ambos os algoritmos concentraram os maiores valores na diagonal principal da matriz de confusão, o que representa assertividade. Os erros foram semelhantes em relação às classes trocadas, exemplo, a classe 0 e 1, classe 2 e 5, porém a assertividade foi maior no Random Forest, o que pode ser melhor visualizado na Tabela 4.1, acurácia do Random Forest foi de 81 contra 65,44 do KNN, vale comentar que olhando as métricas classe a classe o Random Forest é melhor em todos e apenas em um caso (classe 5) ambos modelos tiveram os mesmos resultados.

Classe		Precision	Recall	f1-score	Modelo
dyed-lifted-polyps	81,00	0.76	0.76	0.76	RF
dyed-resection-margins		0.78	0.76	0.77	RF
esophagitis		0.75	0.72	0.73	RF
normal-cecum		0.83	0.95	0.89	RF
normal-pylorus		0.93	0.94	0.93	RF
normal-z-line		0.72	0.77	0.74	RF
polyps		0.85	0.74	0.79	RF
ulcerative-colitis		0.88	0.84	0.86	RF
dyed-lifted-polyps	65,44	0.62	0.30	0.41	KNN
dyed-resection-margins		0.87	0.20	0.33	KNN
esophagitis		0.77	0.56	0.65	KNN
normal-cecum		0.53	0.96	0.68	KNN
normal-pylorus		0.93	0.91	0.92	KNN
normal-z-line		0.63	0.84	0.72	KNN
polyps		0.48	0.57	0.52	KNN
ulcerative-colitis		0.72	0.88	0.79	KNN

Tabela 1 – Resultados MobileNet

4.2 VGG19

Via Imagem 14, é possível perceber que no Random forest os valores se concentraram na diagonal principal da matriz de confusão, o que representa assertividade, ja o KNN isso ocorreu a partir da classe 02. Já na Tabela 4.2, nota-se que mesmo classe a classe o random forest foi superior ao KNN em todos os casos, e de forma geral obteve uma acurácia de 80,31 contra 63,19.

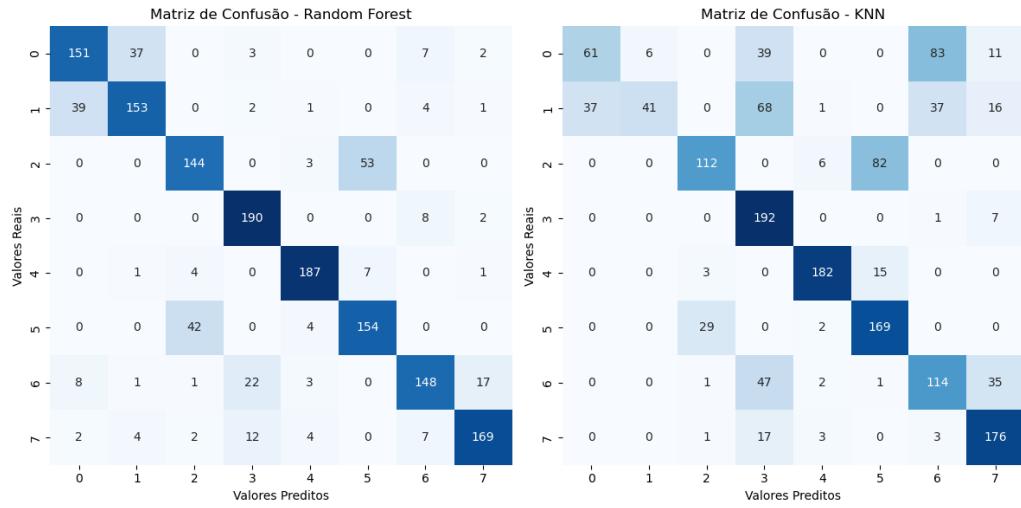


Figura 13 – Matriz de Confusão MobileNet

Classe	Precision	Recall	f1-score	Modelo
dyed-lifted-polyps	0.72	0.61	0.66	RF
dyed-resection-margins	0.71	0.74	0.72	RF
esophagitis	0.77	0.74	0.75	RF
normal-cecum	0.79	0.95	0.86	RF
normal-pylorus	0.95	0.98	0.96	RF
normal-z-line	0.75	0.82	0.78	RF
polyps	0.82	0.73	0.77	RF
ulcerative-colitis	0.90	0.88	0.89	RF
dyed-lifted-polyps	0.71	0.28	0.40	KNN
dyed-resection-margins	0.73	0.37	0.49	KNN
esophagitis	0.57	0.56	0.57	KNN
normal-cecum	0.50	0.96	0.66	KNN
normal-pylorus	0.85	0.91	0.88	KNN
normal-z-line	0.55	0.89	0.68	KNN
polyps	0.80	0.32	0.46	KNN
ulcerative-colitis	0.68	0.77	0.72	KNN

Tabela 2 – Resultados VGG-19

4.3 ResNet50

Via Imagem 15, é possível perceber que ambos os algoritmos concentraram os maiores valores na diagonal principal da matriz de confusão, o que representa assertividade, os erros foram semelhantes em relação às classes trocadas, exemplo, a classe 0 e 1, classe 2 e 5, porém a assertividade foi maior no Random Forest, o que pode ser melhor visualizado na Tabela 4.3, acurácia do Random Forest foi de 81 contra 65,44 do KNN, vale comentar que olhando as métricas classe a classe o Random Forest é melhor em todos e apenas em

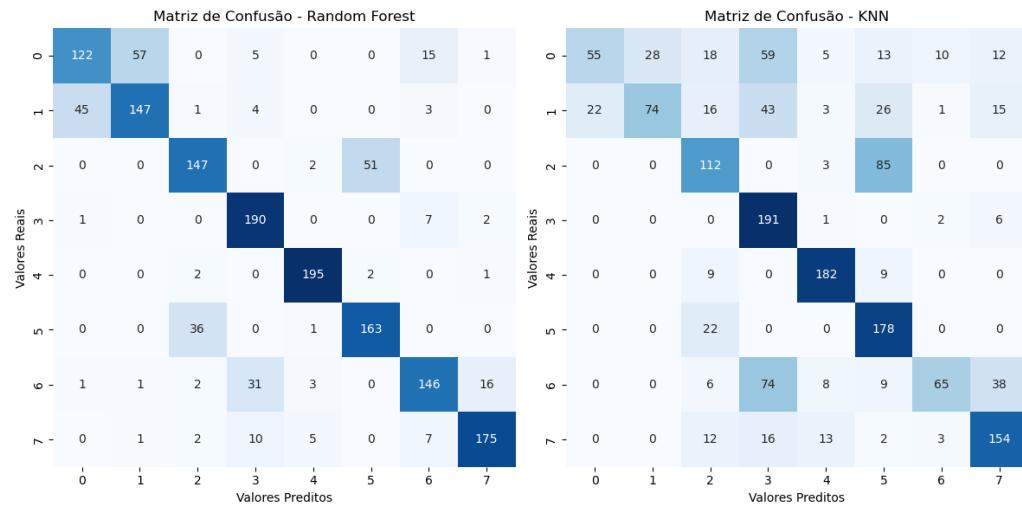


Figura 14 – Matriz de Confusão VGG-19

um casa (classe 5) ambos modelos tiveram os mesmos resultados.

Classe	Precision	Recall	f1-score	Modelo	
dyed-lifted-polyps	0.68	0.68	0.68	RF	
dyed-resection-margins	0.78	0.73	0.75	RF	
esophagitis	0.71	0.71	0.71	RF	
normal-cecum	0.84	0.95	0.89	RF	
normal-pylorus	77,94	0.89	0.93	RF	
normal-z-line		0.72	0.72	RF	
polyps	0.77	0.72	0.75	RF	
ulcerative-colitis	0.83	0.81	0.82	RF	
dyed-lifted-polyps	0.69	0.48	0.57	KNN	
dyed-resection-margins	0.73	0.62	0.67	KNN	
esophagitis	0.67	0.51	0.58	KNN	
normal-cecum	69,69	0.65	0.90	0.76	KNN
normal-pylorus		0.79	0.94	0.86	KNN
normal-z-line		0.63	0.77	0.69	KNN
polyps	0.68	0.66	0.67	KNN	
ulcerative-colitis	0.73	0.73	0.73	KNN	

Tabela 3 – Resultados ResNet50

4.4 Comparação de Modelos

Entre as métricas da matriz de confusão utilizadas para avaliar os modelos, o Random Forest demonstrou ser o mais eficaz. Considerando as redes neurais utilizadas como entrada de dados, foi conduzido um teste adicional para investigar se a assertividade do modelo Random Forest depende ou não dessas redes neurais. Conforme apresentado na Tabela 4, o p-valor obtido foi estatisticamente significativo, indicando que a assertividade

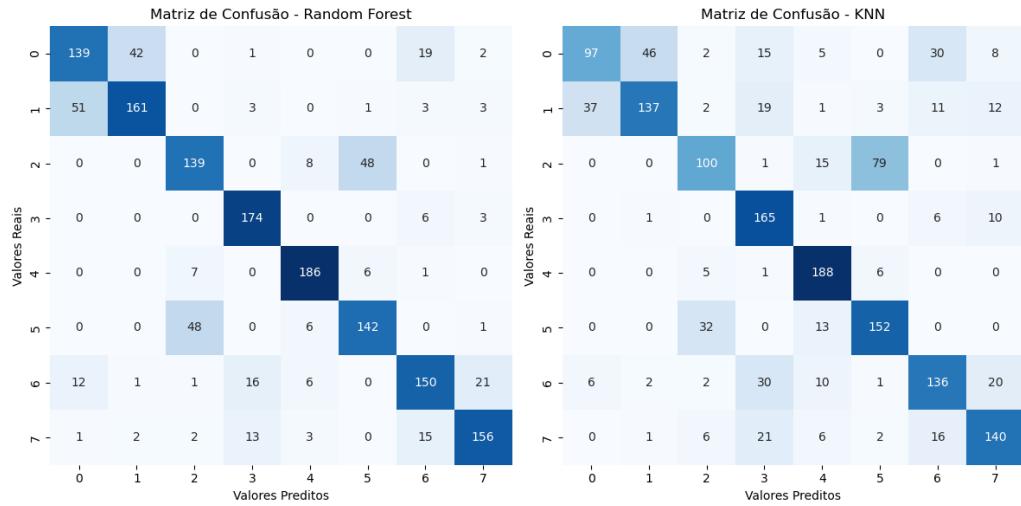


Figura 15 – Matriz de Confusão ResNet50

do modelo depende de algum aspecto relacionado aos modelos neurais. Na análise de correspondência (Anacor) presente na Tabela 5, são apresentadas as associações entre os modelos de redes neurais e os acertos e erros. Observa-se que o modelo MobileNet possui uma associação significativa com os acertos (valores acima de 1,96), enquanto o modelo ResNet50 está associado de maneira significativa aos erros. Em geral, isso sugere que, no caso do MobileNet, os acertos estão acima do esperado, enquanto para o ResNet50, os erros estão acima do esperado. Não foi observada uma associação significativa para a VGG19, indicando que seus valores encontrados estão dentro do esperado quando comparados às outras redes neurais.

Modelo	Acertou		p-valor
	Sim	Não	
MobileNet	1296	247	
VGG19	1285	315	<0,000
ResNet50	1247	353	

Tabela 4 – Teste qui quadrado entre modelos

Modelo	Acertos	Erros
MobileNet	3,97	-3,97
VGG19	-0,49	0,49
ResNet50	-3,45	3,45

Tabela 5 – Associação via ANACOR

5 CONCLUSÕES

O presente estudo explorou a aplicação de redes neurais convolucionais (CNNs) para extração de características em imagens médicas, seguida da utilização de algoritmos de aprendizado de máquina para classificação. Três arquiteturas de CNN foram empregadas: ResNet50, VGG19 e MobileNet, com o intuito de investigar a influência dessas redes na performance dos modelos de aprendizado de máquina.

Os resultados indicam que o processo de Feature Extraction por meio das redes neurais trouxe diferenças substanciais em termos de volume de características, tempo de processamento e desempenho subsequente dos modelos de aprendizado de máquina. O algoritmo Random Forest demonstrou consistentemente melhores métricas de precisão, recall e pontuação F1 em comparação com o KNN para todos os conjuntos de dados gerados pelas CNNs.

Além disso, a análise comparativa entre as redes neurais revelou que o desempenho dos modelos de aprendizado de máquina parece ser sensível à arquitetura da rede neural utilizada na etapa de extração de características. Foi observada uma associação significativa entre a escolha da rede neural e a distribuição dos acertos e erros nos modelos finais, evidenciando que diferentes arquiteturas influenciaram a assertividade dos algoritmos de maneira distinta, onde a arquitetura da MobileNet obteve os melhores valores gerais, MobileNet também foi o mais performático.

Portanto, conclui-se que a escolha da arquitetura de rede neural desempenha um papel crucial na eficácia dos modelos de aprendizado de máquina para tarefas de classificação de imagens. Esta conclusão ressalta a importância de considerações específicas sobre a seleção de redes neurais ao desenvolver sistemas de visão computacional ou tarefas de reconhecimento de imagem, destacando a necessidade de estudos mais aprofundados para compreender como diferentes arquiteturas influenciam diretamente a eficiência dos modelos de aprendizado de máquina em cenários de processamento de imagens médicas.

Para próximos passos poderia ser implementado o modelo em uma aplicação web para classificação automática das imagens, auxiliando o profissional afim de classificar as imagens. Testar novos modelos com processamento paralelo para melhorar a performance e testar outros modelos de machine learning (SVM, XGBoost, entre outros) para tentar melhorar a assertividade dos resultados.

REFERÊNCIAS

- ACADEMY, D. S. **Capítulo 1 – Deep Learning e a Tempestade Perfeita.** 2022. Available at: <<https://www.deeplearningbook.com.br/deep-learning-a-tempestade-perfeita/>>.
- ARAUJO, F. M. d. **A inteligência artificial e os seus impactos no mundo do trabalho.** 2020. Tese (Doutorado), 2020.
- ATKESON, C. G.; MOORE, A. W.; SCHAAAL, S. Locally weighted learning. **Lazy learning**, Springer, p. 11–73, 1997.
- BALLARD, D. H. Computer vision. **Prentice-Hall**, p. 258–259, 1982.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001.
- CHAN, J. S. H. *et al.* Gastrointestinal disease burden and mortality: A public hospital-based study from 2005 to 2014. **Journal of gastroenterology and hepatology**, Wiley Online Library, v. 34, n. 1, p. 124–131, 2019.
- HART, P. E.; STORK, D. G.; DUDA, R. O. **Pattern classification.** [S.l.: s.n.]: Wiley Hoboken, 2000.
- HE, K. *et al.* Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition.** [S.l.: s.n.], 2016. p. 770–778.
- HOWARD, A. G. *et al.* Mobilenets: Efficient convolutional neural networks for mobile vision applications. **arXiv preprint arXiv:1704.04861**, 2017.
- HU, Z. *et al.* Deep learning for image-based cancer detection and diagnosis- a survey. **Pattern Recognition**, Elsevier, v. 83, p. 134–149, 2018.
- JR, J. C. M. d. S. Síndrome do ceco móvel e as doenças gastrintestinais funcionais. **Revista Brasileira de Coloproctologia**, SciELO Brasil, v. 26, p. 463–469, 2006.
- KAUL, V.; ENSLIN, S.; GROSS, S. A. History of artificial intelligence in medicine. **Gastrointestinal endoscopy**, Elsevier, v. 92, n. 4, p. 807–812, 2020.
- LEE, K.-F. **Inteligência artificial.** [S.l.: s.n.]: Globo Livros, 2019.
- LIMA, D. L. S. *et al.* Classificação de imagens de exames de endoscopia por cápsula utilizando transformers. Universidade Federal do Maranhão, 2023.
- MAGALHÃES, T. G. N. L. de. Pólipos gástricos: relevância clínica. 2019.
- MILANO, D. de; HONORATO, L. B. Visão computacional. **Faculdade de Tecnologia, Universidade Estadual de Campinas**, 2010.
- MONGELO, A. I. *et al.* Validação de método baseado em visão computacional para automação da contagem de viabilidade de leveduras em indústrias alcooleiras. **Universidade Católica Dom Bosco, Campo Grande, MG**, 2012.

POGORELOV, K. *et al.* Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In: **Proceedings of the 8th ACM on Multimedia Systems Conference**. New York, NY, USA: ACM, 2017. (MMSys'17), p. 164–169. ISBN 978-1-4503-5002-0.

SIEGEL, S. **Estatística não-paramétrica**. São Paulo: Ed. [S.l.: s.n.]: McGraw-Hill do Brasil, 1981.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

THOMAZ, V. de A. **Avaliação de Aumento de Dados via Geração de Imagens Sintéticas para Segmentação e Detecção de Pólipos em Imagens de Colonoscopia Utilizando Aprendizado de Máquina**. 2020. Tese (Doutorado) — PUC-Rio, 2020.

WIKIDOCS. **K₀4Understanding of MobileNet**. 2023. Available at : <>.

WIKIDOCS. **K₀4Understanding of VGG – 16, VGG – 19**. 2023. Available at : <>.

WIKIDOCS. **K₀5Understanding of ResNet**. 2023. Available at : <>.