

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Vitor Barbosa Lemes Fernandes

**Compressão de Redes Neurais com  
Decomposição de Tucker para Classificação de  
Imagens Histológicas de Displasia**

**Uberlândia, Brasil**

**2023**

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Vitor Barbosa Lemes Fernandes

**Compressão de Redes Neurais com Decomposição de Tucker para Classificação de Imagens Histológicas de Displasia**

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Orientador: Marcelo Zanchetta do Nascimento

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2023


**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**

Faculdade de Computação

Av. João Naves de Ávila, nº 2121, Bloco 1A - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902

 Telefone: (34) 3239-4144 - <http://www.portal.facom.ufu.br/> facom@ufu.br

**ATA DE DEFESA - GRADUAÇÃO**

Curso de Graduação em:	Bacharelado em Ciência de Computação				
Defesa de:	Projeto de Graduação 2				
Data:	05/12/2023	Hora de início:	15:00	Hora de encerramento:	15:45
Matrícula do Discente:	11921BCC035				
Nome do Discente:	Vitor Barbosa Lemes Fernandes				
Título do Trabalho:	Compressão de Redes Neurais com Decomposição de Tucker para Classificação de Imagens Histológicas de Displasia				
A carga horária curricular foi cumprida integralmente?	<input checked="" type="checkbox"/> Sim ( ) Não				

Reuniu-se na Sala virtual da plataforma MS Teams, da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Curso de Graduação em Ciência da Computação, assim composta: - Professores Henrique Coelho Fernandes - FACOM/UFU, Paulo Henrique Ribeiro Gabriel - FACOM/UFU; e Marcelo Zanchetta do Nascimento - FACOM/UFU, orientador do candidato.

Iniciando os trabalhos, o presidente da mesa, Dr. Marcelo Zanchetta do Nascimento, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao discente a palavra, para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do curso.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

( X ) Aprovado(a) Nota [100]

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Marcelo Zanchetta do Nascimento, Professor(a) do Magistério Superior**, em 05/12/2023, às 18:24, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Paulo Henrique Ribeiro Gabriel, Professor(a) do Magistério Superior**, em 05/12/2023, às 18:42, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Henrique Coelho Fernandes, Professor(a) do Magistério Superior**, em 05/12/2023, às 18:58, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).

---



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **5016897** e o código CRC **0FAE04D0**.

---

# Agradecimentos

Agradeço a todos os colegas e professores que contribuíram com a minha jornada, em especial ao professor Marcelo Zanchetta, por toda a paciência e prestatividade em sua orientação ao longo deste trabalho.

Agradeço também minha família. Meu pais, que sempre me deram o incentivo necessário para continuar e sempre apoiaram as minhas decisões, mesmo quando a decisão era desistir. Minha irmã Laura, que esporadicamente tornou os dias de escrita deste trabalho menos solitários. E minha avó Helena, que me disse que tudo daria certo, mesmo sem muita compreensão do que se tratava.

Por fim, agradeço a todos os outros que contribuíram, direta ou indiretamente, no desenvolvimento deste trabalho.

# Resumo

O câncer na cavidade oral é um dos tipos mais comuns, tornando necessário para um diagnóstico precoce a investigação de lesões com potencial de desenvolvimento para a doença. Displasias são lesões de estágio inicial, que podem se desenvolver para estágios mais severos, e são caracterizadas por alterações no formato e tamanho do núcleo das células do tecido epitelial. A classificação manual de imagens histológicas é um processo demorado e sua qualidade depende da experiência do especialista. Nos últimos anos os Sistemas de Auxílio ao Diagnóstico (do inglês, *computer aided diagnosis* - CAD) se tornaram aliados no processo de classificação dessas imagens, devido aos avanços nas áreas de processamento de imagens e inteligência artificial. Apesar disso, técnicas empregadas em processos de um sistema CAD, como as redes neurais convolucionais, frequentemente sofrem de problemas de super-parametrização. Este trabalho apresenta uma investigação do uso da decomposição de Tucker nos tensores do *kernel* das camadas convolucionais de arquitetura ResNet para a compressão do número de parâmetros. Além de avaliar a classificação com as redes, foram avaliados os descritores em camadas da rede, denominados mapas de características, com os algoritmos de aprendizado de máquina. Os resultados mostram que a decomposição consegue uma diminuição no número de parâmetros, mantendo uma acurácia relevante, especialmente quando utilizados com as redes para a classificação de dados em bancos de imagem balanceados.

**Palavras-chave:** redes convolucionais, histologia, classificação, compressão, tensores, Tucker.

# Lista de ilustrações

Figura 1 – Comparação entre tecidos saudável e com displasia . . . . .	14
Figura 2 – Etapas de um sistema de auxílio ao diagnóstico. . . . .	16
Figura 3 – Exemplo de uma rede neural artificial com as conexões e neurônios. . . . .	18
Figura 4 – Estrutura clássica de um neurônio artificial. . . . .	19
Figura 5 – Estrutura de uma rede neural convolucional. . . . .	20
Figura 6 – Comparação entre diferentes técnicas de <i>pooling</i> . . . . .	21
Figura 7 – Exemplo simples de uma Árvore de Decisão. . . . .	22
Figura 8 – Etapas dos métodos e materiais empregados para classificação das lesões nas imagens histológicas. . . . .	30
Figura 9 – Exemplos de imagens do banco elaborado por Silva. . . . .	31
Figura 10 – Exemplos de imagens do banco elaborado por Costa. . . . .	32
Figura 11 – Exemplos de transformações para aumento de dados. . . . .	33
Figura 12 – Arquitetura de uma ResNet-50. . . . .	34
Figura 13 – A decomposição de Tucker aplicada em um tensor. . . . .	35
Figura 14 – Comparação entre os resultados obtidos nos algoritmos de ML avaliando o Banco Silva Balanceado. . . . .	42
Figura 15 – Comparação entre os resultados obtidos nos algoritmos de ML avaliando o Banco Silva Desbalanceado. . . . .	43
Figura 16 – Comparação entre os resultados obtidos nos algoritmos de ML avaliando o Banco Costa Balanceado. . . . .	44
Figura 17 – Comparação entre os resultados obtidos nos algoritmos de ML avaliando o Banco Costa Desbalanceado. . . . .	45

# Lista de tabelas

Tabela 1 – Número de imagens de cada classe presentes em cada banco de imagem.	32
Tabela 2 – Parâmetros dos diferentes algoritmos classificadores implementados em scikit-learn . . . . .	36
Tabela 3 – Valores obtidos com as redes convolucionais decompostas avaliando o banco de imagens Silva balanceado. . . . .	37
Tabela 4 – Valores obtidos com as redes convolucionais decompostas avaliando o banco de imagens Silva desbalanceado. . . . .	38
Tabela 5 – Valores obtidos com as redes convolucionais decompostas avaliando o banco de imagens Costa balanceado. . . . .	39
Tabela 6 – Valores obtidos com as redes convolucionais decompostas avaliando o banco de imagens Costa desbalanceado. . . . .	40
Tabela 7 – Análise entre métodos de classificação de lesões na cavidade oral propostos na literatura. . . . .	43

# Lista de abreviaturas e siglas

RN	Redes Neurais
CNN	Convolutional Neural Network
CAD	Computer Aided Diagnosis
ML	Machine Learning
DT	Decision Tree
RF	Random Forest
SVM	Support Vector Machine
NB	Naive Bayes

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>10</b>
1.1	Justificativa	11
1.2	Objetivos	11
1.3	Organização desse Trabalho de Conclusão de Curso	12
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>13</b>
2.1	Conceitos Fundamentais	13
2.1.1	Displasia	13
2.1.2	Sistema de Diagnóstico Auxiliado por Computador	14
2.1.3	Redes Neurais Artificiais	17
2.1.4	Redes Neurais Convolucionais	19
2.1.5	Decomposição de Tensores	21
2.1.6	Algoritmos de Aprendizado de Máquina	22
2.1.7	Métricas de Avaliação de Classificação	23
2.2	Estado da Arte	24
2.2.1	Classificação de Displasia Auxiliada por Computador	24
2.2.2	Redes Convolucionais com Decomposição de Tensores	26
2.2.3	Considerações Finais	29
<b>3</b>	<b>MATERIAIS E MÉTODOS</b>	<b>30</b>
3.1	Banco de imagens	30
3.2	Aumento de Dados	32
3.3	Arquitetura ResNet	33
3.4	Decomposição de Tucker	34
3.5	Avaliação de Desempenho das Abordagens	35
<b>4</b>	<b>RESULTADOS</b>	<b>37</b>
4.1	Classificação com os modelos ResNet com Banco de Imagens Silva	37
4.2	Resultados no Banco de Imagens Costa com a ResNet	39
4.3	Classificação com os Algoritmos de ML	41
4.4	Análise de Desempenho em Relação aos Estudos	42
<b>5</b>	<b>CONCLUSÃO</b>	<b>46</b>
	<b>REFERÊNCIAS</b>	<b>47</b>

# 1 Introdução

As displasias são lesões pré-invasivas que prejudicam o desenvolvimento das células de órgãos ou tecidos e podem ser detectadas por meio de características morfológicas como o crescimento anormal dessas células, diferenças em seu formato, hipercromasia dentre outras características (KUMAR, 2010). Segundo INCA (2002), a detecção dessas lesões de forma prematura é importante pois, quando ainda em estágio inicial, as displasias podem ser revertidas antes de evoluírem para graus mais severos. Entretanto, ao atingirem graus mais avançados de desenvolvimento, estas lesões tendem a evoluir para um câncer, quando a sua replicação se torna fora de controle e suas marcações morfológicas se tornam ainda mais exacerbadas que no grau severo.

Como mencionando por (CHEN; CHEN, 2022), fazer a classificação manual de imagens histológicas é um processo demorado e tedioso. Além disso, a qualidade e coerência da classificação varia de acordo com a experiência dos profissionais encarregados, sendo que profissionais mais experientes conseguem entregar classificações mais precisas e que tendem a concordar com classificações feitas por outros profissionais com o mesmo nível de experiência (WARNAKULASURIYA et al., 2008). Pensando nessas dificuldades, os sistemas de auxílio ao diagnóstico (do inglês, *computer aided diagnosis* - CAD) tem como seu objetivo ajudar profissionais da saúde a classificarem imagens histológicas de forma mais rápida e precisa, evitando o diagnóstico errado e tornando o processo menos demorado e trabalhoso. O uso de inteligência artificial em aplicações médicas foi proposto desde os primórdios da computação, porém somente nos últimos anos é que este tipo de sistema tem sido utilizado de maneira significativa para o auxílio ao diagnóstico (LISBOA, 2002). Desde então, diversas técnicas vêm sendo aplicadas como CAD, como árvores de decisão, regressão logística, florestas aleatórias, rede neurais artificiais, dentre outras, sempre consistindo de métodos que utilizam modelos matemáticos para fazer suas previsões (PAIXÃO et al., 2022).

Apesar de relativamente eficientes, os CAD atuais possuem características que podem ser melhoradas. De acordo com Kim et al. (2016), redes neurais convolucionais (do inglês, *convolutional neural networks* - CNN) podem muitas vezes ser super-parametrizadas, o que significa que possuem um número muito grande de parâmetros, tornando-as muito pesadas e lentas, mesmo na utilização para classificação depois de já treinadas. Ainda esses autores descrevem que a super-parametrização faz sentido na hora de treinar as redes, pois ajuda a função de erro a convergir para seu mínimo global mais rapidamente, porém, após a etapa de treinamento, podemos aplicar técnicas para diminuir o número de parâmetros da rede.

Dado o alto grau de incidência de câncer na cavidade oral, as dificuldades apresentadas acima para o correto diagnóstico e classificação das displasias nesta região e o problema de super-parametrização de redes neurais profundas, este trabalho propõe-se a investigar técnicas de decomposição de tensores que possam ser utilizadas nos núcleos das camadas convolucionais das redes CAD treinadas para classificar imagens histológicas de lesões na cavidade bucal, a fim de diminuir o número de parâmetros dessas redes, deixando-as mais leves, e permitindo um desempenho melhor mesmo em sistemas computacionais mais fracos.

## 1.1 Justificativa

De acordo com [Lima et al. \(2023\)](#), o câncer hoje é uma das principais causas de morte no mundo, sendo o principal problema de saúde pública a nível global, correspondendo a uma das maiores causas de morte prematura na maioria dos países. Somente no Brasil, a estimativa é que entre 2023 e 2025 ocorrerão cerca de 704 mil novos casos de câncer, sendo o câncer da cavidade oral um dos mais comuns, estimando-se que estará entre os dez com maior número de incidência, com cerca de 15.100 novos casos no país. Ainda de acordo com [Lima et al. \(2023\)](#), este tipo de câncer é mais comum em homens, com uma estimativa de 10,30 novos casos para cada 100 mil indivíduos do sexo masculino, contra 3,83 casos a cada 100 mil mulheres. O número de óbitos registrados no Brasil para este tipo de câncer foi de 6.192 no ano de 2020. Dito isto, torna-se evidente a importância do diagnóstico de casos de displasia o mais precocemente possível, a fim de evitar sua evolução para um câncer.

Uma forma de aprimorar a velocidade do diagnóstico destas lesões é utilizando redes neurais para a classificação das imagens histológicas. Como ressaltado por [Kim et al. \(2016\)](#), redes neurais profundas tem a tendência de ser super parametrizadas, entretanto, após treinadas, podem ser comprimidas utilizando técnicas de decomposição de tensores, resultando em uma rede com menos parâmetros, porém sem perder acurácia significativa. Dito isto, o uso da decomposição de Tucker aplicada ao *kernel* de redes neurais convolucionais pode ser uma alternativa interessante para reduzir o número de parâmetros dessas redes, mantendo sua qualidade de classificação, mas tornando-a mais rápida e com um custo computacional mais baixo, melhorando seu desempenho em sistemas que possuem baixo poder computacional.

## 1.2 Objetivos

Este trabalho teve como objetivo investigar técnicas de decomposição de tensores no *kernel* de camadas convolucionais a fim de diminuir o número de parâmetros de redes

neurais utilizadas na classificação de imagens histológicas de lesões na região da cavidade oral. Pontualmente, os objetivos específicos deste trabalho foram:

- Investigação do uso da Decomposição de Tucker em tensores;
- Estudo sobre a aplicação de decomposição no *kernel* de camadas convolucionais para a compressão de redes neurais;
- Avaliação do método sobre os modelos de CNN ResNet;
- Análise de métricas estatísticas para avaliação de desempenho das redes comprimidas sobre imagens histológicas.

### 1.3 Organização desse Trabalho de Conclusão de Curso

Neste capítulo, foram apresentadas as considerações iniciais, motivação e objetivos do desenvolvimento deste trabalho. O restante deste trabalho possui a seguinte organização:

- Capítulo 2: Introdz os principais conceitos para a compreensão e validação do trabalho proposto e apresenta os trabalhos correlatos relacionados com a proposta;
- Capítulo 3: Descreve o trabalho proposto, bem como as técnicas e metodologias desenvolvidas;
- Capítulo 4: Apresenta os resultados alcançados e as respectivas análises do método de classificação, além de um comparativo com a literatura;
- Capítulo 5: São apresentadas as contribuições obtidas, considerações finais e trabalhos futuros.

## 2 Referencial Teórico

Neste capítulo são abordados os conceitos fundamentais para o entendimento do trabalho. Este capítulo também apresenta o tópico de Estado da Arte sobre os temas de decomposição de tensores e sobre classificação de displasia, apresentando os trabalhos correlatos de diferentes autores.

### 2.1 Conceitos Fundamentais

Nesta seção estão uma introdução aos conceitos teóricos sobre redes neurais convolucionais e decomposição de tensores, além de explicações dos sistemas de apoio ao diagnóstico e informações sobre displasia da cavidade oral.

#### 2.1.1 Displasia

As displasias são um dos tipos de lesões que geralmente ocorrem no tecido epitelial, a qual é caracterizada por problemas no desenvolvimento das células, levando as características morfológicas distintas, como formatos de núcleo em relação as células saudáveis e hiper Cromasia, ou seja, tem o excesso de cromatina no protoplasma, crescimento anormal do núcleo, variação anormal no tamanho da célula, dentre outras características (KUMAR, 2010; WARNAKULASURIYA et al., 2008). Na maioria das vezes, as displasias se apresentam como lesões brancas aparecendo na parte inferior da boca ou na língua, mas pode ocorrer como lesões vermelhas ou úlceras, e aparecer em outras partes da cavidade oral, como na gengiva ou nos lábios (LUMERMAN; FREEDMAN; KERPEL, 1995).

Essas lesões tem maior tendencia a aparecer quando a região bucal possui contato com algum tipo de substância carcinógena. A principal causa do aparecimento de displasias é o tabagismo, que pode aumentar a chance de desenvolver câncer em até seis vezes (SADRI; MAHJUB, 2007). A segunda é o consumo constante de bebidas alcoólicas. Segundo dados disponíveis em Atty et al. (2022), pelo menos um dos dois hábitos está presente em 90% dos casos da doença.

De acordo com Warnakulasuriya et al. (2008), os graus de displasia são geralmente divididos em três categorias, sendo elas, leve, moderada e severa. Esses graus descrevem o quanto as características das lesões se tornam proeminentes, sendo que quanto maior o grau mais evidente elas se tornam. De acordo com o INCA (2002), quando detectadas, essas lesões devem ser tratadas o mais rapidamente possível para evitar que evoluam para graus mais avançados, pois uma vez em grau severo torna-se difícil o tratamento para reversão para tecido saudável. Além disso, existe uma grande chance de evolução para um

câncer, onde as características apresentadas na displasia tornam-se ainda mais evidentes e a multiplicação das células com a condição se torna fora de controle.

Na Figura 1 pode-se observar uma comparação entre amostras de tecidos saudáveis e com displasia severa.

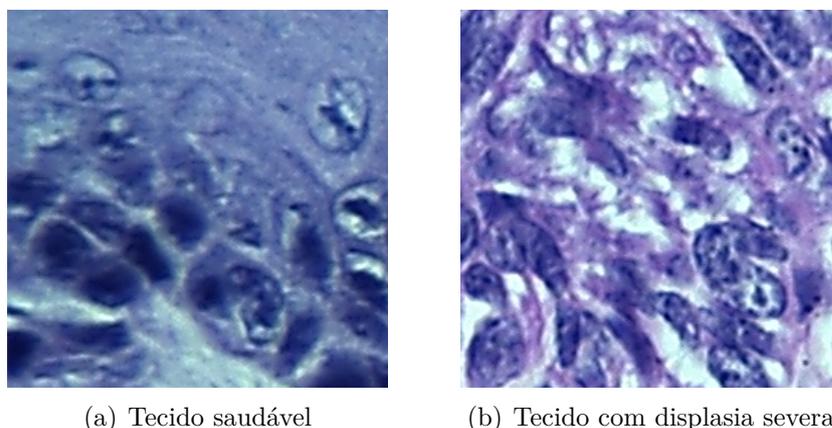


Figura 1 – Comparação entre amostras de tecido saudável e com displasia severa. Imagem extraída de [Silva et al. \(2022a\)](#).

O grau da displasia pode ser classificado de acordo com a definição de [Lumerman, Freedman e Kerpel \(1995\)](#) onde a displasia leve é caracterizada por uma extensão de no máximo um terço do tecido epitelial e por pouca perturbação nas características das células. Já a displasia moderada pode ser classificada pela extensão de mais de dois terços do epitelial, apesar de existirem casos em que uma displasia pode ser considerada leve ou severa mesmo ocupando cerca de dois terços do epitelial, a depender do quanto a estrutura das células está comprometida. Por último, o grau severo de displasia pode ser considerado quando atinge mais de dois terços do tecido epitelial e a estrutura das células apresenta as características clássicas de displasia de maneira bastante acentuada.

Um problema a ser notado é que, de acordo com [Warnakulasuriya et al. \(2008\)](#), os profissionais responsáveis pela classificação nem sempre entram em consenso entre si, existindo características que são mais universais e outras que tendem a ser classificadas de maneira divergente entre diferentes classificadores.

### 2.1.2 Sistema de Diagnóstico Auxiliado por Computador

O aprendizado de máquina (do inglês, *machine learning* - ML) são técnicas computacionais da área da inteligência artificial que são utilizadas para o aprendizado de padrões a partir de um banco de dados com o objetivo de conseguir identificar, classificar e prever novas entradas a partir dos padrões observados nos dados que o algoritmo foi treinado ([PAIXÃO et al., 2022](#)). O uso destas técnicas auxilia diferentes áreas da medicina, como a descobertas de novos remédios, a identificação de pacientes em risco e o diagnóstico de

doenças (KUMAR et al., 2022). Este último caso é denominado de sistema de apoio ao diagnóstico (do inglês, *computer aided diagnosis* - CAD).

Segundo Lisboa (2002), o uso de inteligência artificial para o auxílio em áreas clínicas não é uma ideia nova, e tem sido proposta desde os primórdios da computação, tendo porém se tornado mais popular somente nas últimas décadas, devido ao aumento do poder computacional e a maior disponibilidade de dados para o treinamento destes sistemas. Diferentes tipos de técnicas são utilizadas no CAD, como *Random Forest*, Regressão Logística, *Boltzmann Machine* e Redes Neurais Artificiais, e seu uso é bastante promissor, pois devido ao maior poder computacional e a grande quantidade de dados disponíveis, é possível analisar grandes quantidades de entradas de forma rápida que uma classificação, contribuindo para que o paciente tenha um diagnóstico e possa começar o tratamento o mais breve possível (KUMAR et al., 2022).

De maneira geral, os sistemas CAD podem ser divididos em cinco etapas bem estabelecidas, sendo elas: aquisição, pré-processamento, segmentação, extração de características e classificação (GONZALEZ; WOODS, 2009). Na Figura 2 é ilustrado cada uma destas etapas.

A etapa de aquisição tem a captura da imagem de forma digital e transformação em um arquivo. Neste etapa, estão envolvidas as decisões de escolha de sensores, digitalizadores, lentes, ambiente, temperatura dentre outras características para que a imagem possa ser capturada de maneira eficiente e convertida em um formato digital, para então passar pelas outras etapas do sistema CAD (GONZALEZ; WOODS, 2009). Alguns autores, como Kaushal et al. (2019), não consideram a aquisição das imagens como parte do sistema CAD, considerando a etapa de pós-processamento como a primeira etapa do processo.

A etapa de pré-processamento se destina a tratar possíveis imperfeições que as imagens digitalizadas sofreu durante a aquisição. Durante a etapa de aquisição, dependendo do sensor e do digitalizador, pode surgir inconsciências como variações nas cores, no contraste e ruído nas imagens. Isso pode interferir diretamente nas etapas seguintes de segmentação e extração de características. Por isso, nessa etapa são empregados diferentes técnicas, como normalização de cores ou corantes, redução de ruído, as quais tentam atenuar as imperfeições (KAUSHAL et al., 2019).

A etapa de segmentação é a responsável por separar os objetos de interesse da imagem, e pode ser feita utilizando técnicas de clusterização, detecção de bordas e detecção de regiões (KAUSHAL et al., 2019). Essa etapa é considerada difícil de execução e a escolha do algoritmo que deve garantir resultados promissores (GONZALEZ; WOODS, 2009). Parte das aproximações desenvolvidas com essas técnicas empregadas nessa etapa não podem ser generalizadas para outros domínios. Isso torna a etapa mais complexa durante o desenvolvimento das aproximações. Já etapa de extração de características consiste em

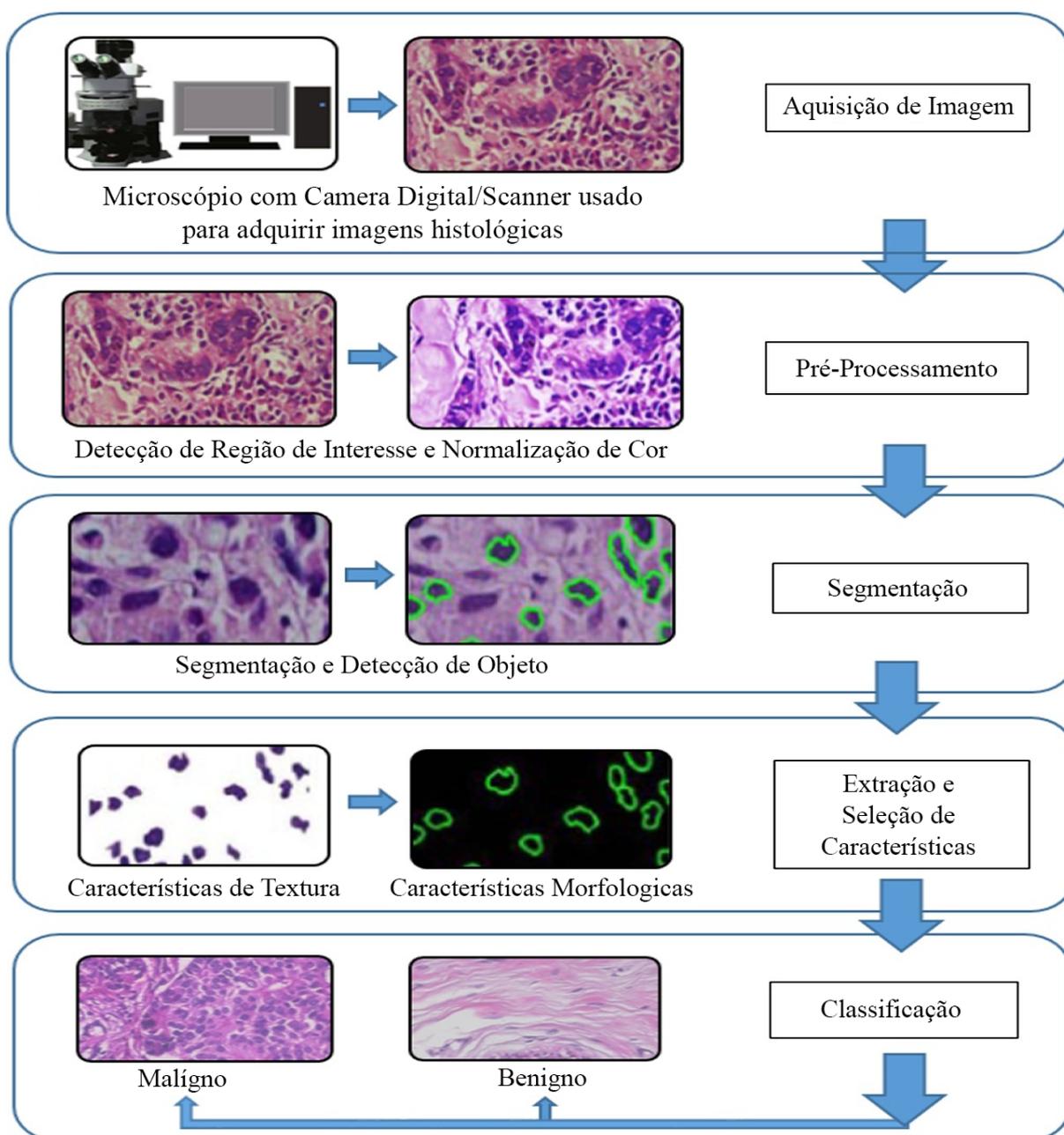


Figura 2 – As cinco etapas de um sistema de auxílio ao diagnóstico. Imagem adaptada de [Kaushal et al. \(2019\)](#)

extrair descritores importantes das imagens resultantes após a etapa de segmentação. A extração pode ser feita baseada em características morfológicas, de textura ou intensidade, dentre outras formas comuns de se definir áreas distintas, seja a nível de tecido ou a nível celular ([KAUSHAL et al., 2019](#)).

Por último, na etapa de classificação, é onde as imagens obtidas pelas etapas anteriores são classificadas em categorias com base no problema que se está resolvendo. Esta classificação pode ser feita utilizando diferentes algoritmos, como regressão logística, K-Nearest Neighborhood(KNN) ou redes neurais ([KAUSHAL et al., 2019](#)).

### 2.1.3 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) são técnicas computacionais que tem como inspiração as redes neurais biológicas. De acordo com [Oliveira \(2018\)](#), esse tipo de técnica surgiu quando os cientistas da computação passaram a buscar um tipo de algoritmo que pudesse sair dos modelos de binaridade dos algoritmos convencionais da época e adotasse uma estrutura que se assemelhasse a como funciona o cérebro humano, levando ao que seria uma das mais revolucionárias áreas da inteligência artificial. Segundo autor em [Haykin \(2009\)](#), as RNA usa o princípio do cérebro humano, tendo como características: i) tem a capacidade de adquirir conhecimento a partir de um processo de aprendizado; ii) têm seus neurônios artificiais conectados a partir de pesos sinápticos, que possuem a capacidade de guardar o que foi aprendido e transmitir os sinais de acordo com a relevância atribuída a esses pesos.

Uma rede neural tradicionalmente é dividida entre camada de entrada, camadas ocultas e camada de saída. A camada de entrada é a camada que recebe os dados, que dependendo da aplicação podem não ser digitais, devendo então ser convertidos por esta primeira camada antes de serem repassados para a próxima. A camada de saída é a última camada da rede e é nela que se encontra o resultado da entrada, após ser processada por toda a rede. As camadas ocultas são opcionais e no geral quanto maior sua quantidade, maior a capacidade de aprendizado e generalização da rede, porém o custo computacional também aumenta ([OLIVEIRA, 2018](#)).

Na Figura 3 pode-se observar o que seria a representação de uma rede neural artificial com duas camadas ocultas. Cada círculo representa um neurônio artificial e as arestas ligando um neurônio ao outro representam as sinapses que contém os pesos da rede.

Toda a base das redes neurais é formada pelos neurônios, que são as unidades que processam a informação através da rede e formam as diversas camadas de uma RNA. A estrutura clássica de um neurônio pode ser observada na Figura 4. Segundo [Haykin \(2009\)](#), os neurônios podem ser caracterizados a partir de três elementos básicos:

- uma série de pesos, um para cada conexão entre dois neurônios, que são responsáveis por multiplicar a entrada recebida com o objetivo de alterar seu valor de modo a levar a saída esperada;
- um somador responsável por somar os sinais já multiplicados pelos pesos, funcionando como um combinador linear;
- uma função de ativação, que tem o objetivo de transformar os valores obtidos na entrada do neurônio antes de encaminhá-los a saída;

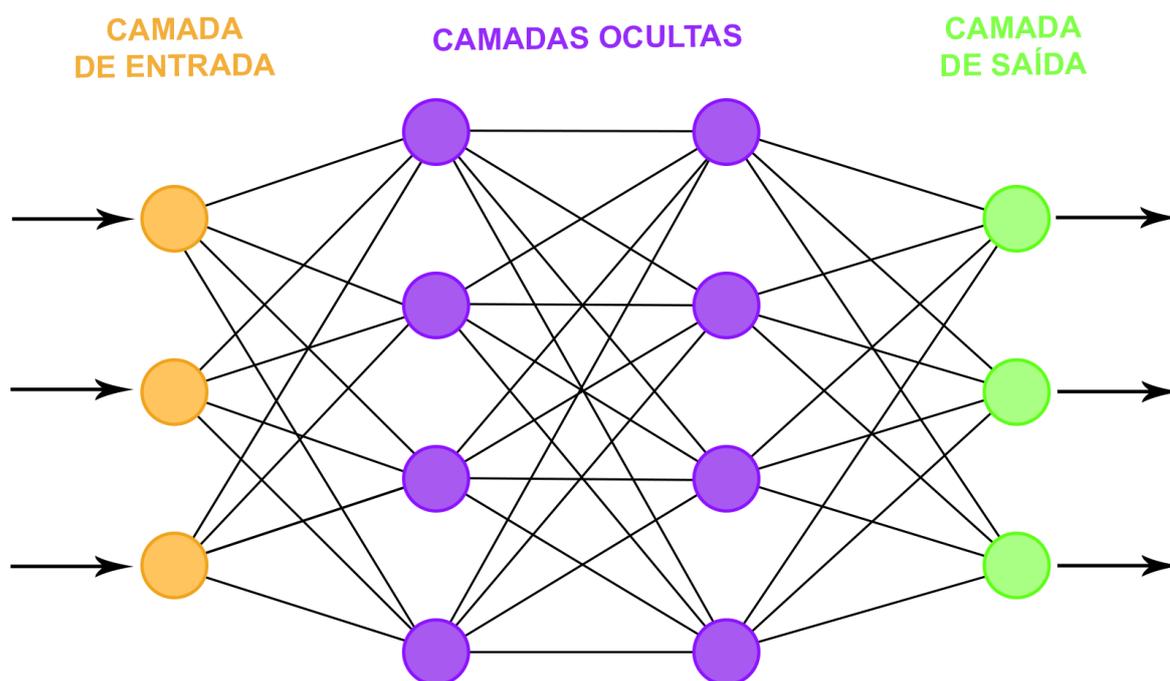


Figura 3 – Exemplo de uma rede neural artificial com as conexões e neurônios.

Além dos pesos, um neurônio pode vir a ter um viés (bias) que é adicionado ao valor calculado após a passagem pelas sinapses e antes da passagem pela função de ativação e tem como objetivo modificar o valor atual para que seja maior ou menor, levando em consideração o objetivo do estudo (HAYKIN, 1999).

Sobre as funções de ativação, segundo Haykin (2009) uma RNA pode ter diferentes tipos de funções, variando de funções muito simples como a clássica função de *Threshold*, que só possui como saída o valor 0 ou 1, até funções mais complexas, como a função Sigmoide, uma função em forma de *S*, não linear e que pode ter a intensidade de sua curva alterada a partir do ajuste de um parâmetro, sendo uma das funções mais utilizadas para redes neurais hoje em dia.

Para trabalhar com uma RNA, são explorado as etapas: i) treinamento; ii) teste. Na etapa de treinamento é onde a RNA recebe entradas, as processa ao longo de suas camadas e compara suas classes com a a probabilidade gerada como saída. A RNA então utiliza um mecanismo chamado *backpropagation* para atualizar os pesos e vieses dos neurônios de modo que quando as entradas forem reinseridas elas resultem em uma saída mais próxima da esperada. Esse processo é repetido várias vezes, até que a taxa de erro seja pequena. Na etapa de teste, a rede não atualiza mais os pesos e simplesmente avalia as entradas que recebe com um novo conjunto de dados denominado grupo de teste (OLIVEIRA, 2018).

O que geralmente se espera após o treinamento de uma RNA é a boa capacidade de generalização. A generalização define a aptidão de uma rede em conseguir produzir

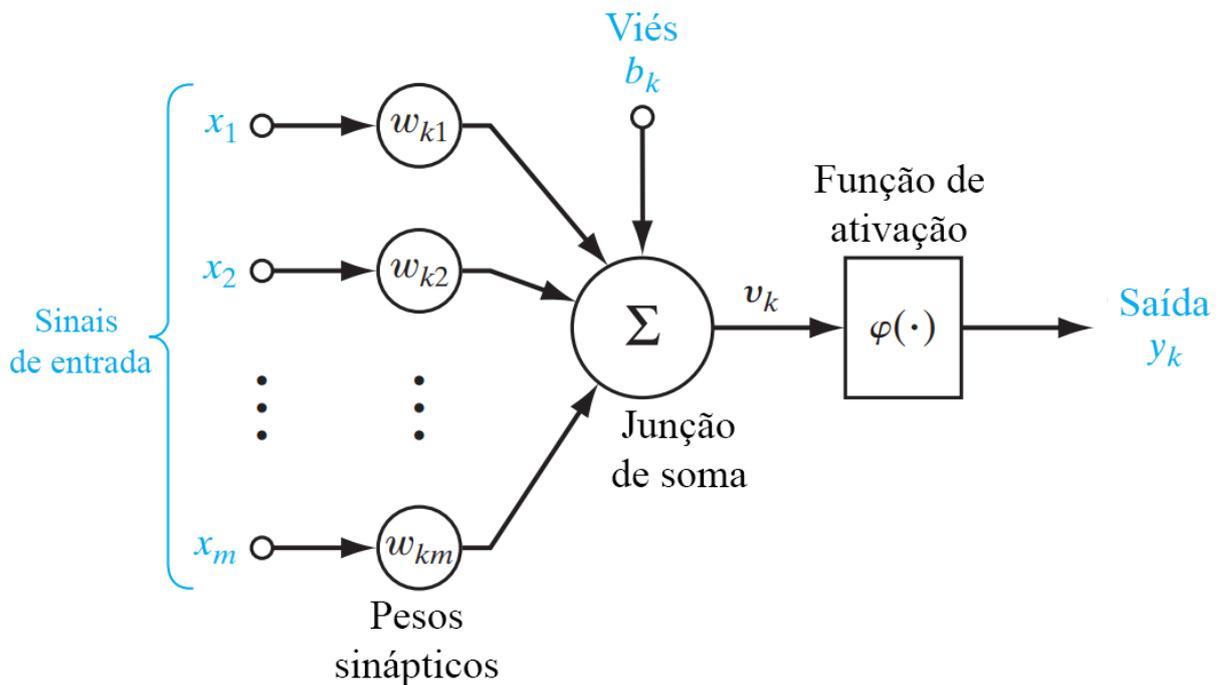


Figura 4 – Estrutura clássica de um neurônio artificial. Imagem adaptada de Haykin (2009)

saídas corretas a partir de entradas não conhecida, usando apenas o conhecimento que foi adquirido a partir das entradas do treinamento (HAYKIN, 1999). Diversos modelos de RNA são apresentados na literatura como perceptrons multicamadas, as convolucionais, recorrente e *auto encoders*. Em imagens digitais, para um processo abrangente nas etapas de um CAD, os modelos baseados em operações convolucionais são os mais empregados na literatura.

#### 2.1.4 Redes Neurais Convolucionais

Propostas por Lecun et al. (1998), as redes neurais convolucionais (do inglês, *convolutional neural networks* - CNN) são redes neurais que utilizam em algumas de suas camadas operações de convolução ao invés da multiplicação de matrizes das redes neurais tradicionais. As CNNs são mais comumente utilizadas para reconhecimento e classificação de padrões em imagens, que podem ser representadas com cada canal de cor sendo uma matriz 2D, e cada entrada da matriz representando um pixel (GOODFELLOW; BENGIO; COURVILLE, 2016). A Figura 5 apresenta uma representação da estrutura básica de uma rede convolucional plana.

De acordo com Haykin (2009), as CNNs são especialmente desenvolvidas para reconhecer e aprender padrões de forma, contorno e bordas, mesmo que essas características sejam apresentadas com modificações de rotação, posição, inclinação e escala de uma imagem para outra. Para conseguir fazer isso, as CNNs utilizam dos seguintes recursos: (i)

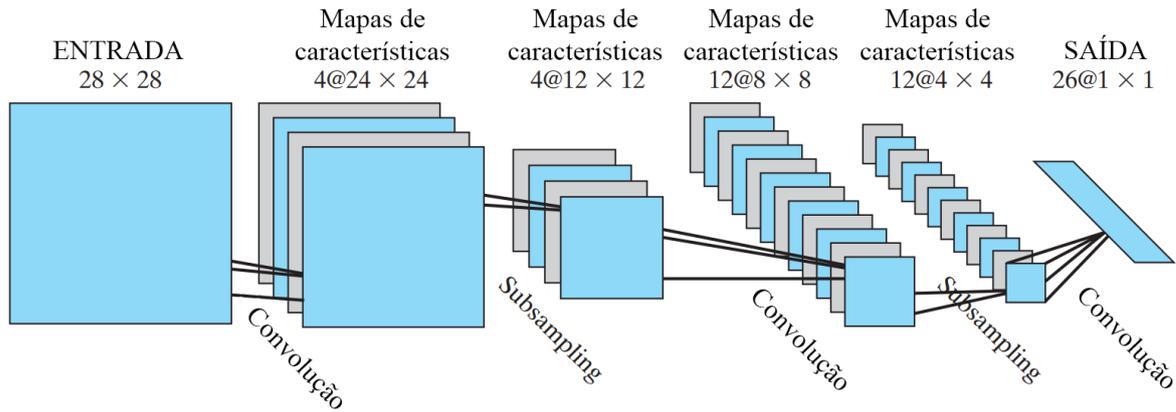


Figura 5 – Estrutura de uma rede convolucional. Imagem adaptada de Haykin (2009)

extração de características; (ii) mapa de características; (iii) *subsampling*.

A extração de características acontece porque os neurônios de cada camada recebem as sinapses de entrada dos neurônios da camada anterior a partir de um campo receptivo local que se desloca ao longo da imagem, aprendendo características da entrada. Com essas características, a camada consegue identificá-las em outras imagens mesmo que elas não estejam exatamente na mesma posição que aprendeu, desde que o contexto de outras características se mantenha semelhante. Os mapas de características são os resultados da extração de características. Cada camada convolucional irá gerar vários mapas e eles são compostos pelos pesos sinápticos que são compartilhados por vários neurônios da camada. O número de mapas presente em cada camada é definido pelo usuário durante a configuração da CNN. Por último, o *subsampling* é onde a dimensão dos mapas de características é diminuída, com a intenção de tornar menos relevante o local ou formato exato de uma característica extraída, permitindo que a capacidade de generalização da rede melhore (HAYKIN, 2009).

De acordo com Goodfellow, Bengio e Courville (2016), assim como nas camadas tradicionais, após passar pelas operações envolvendo os pesos do neurônio e o viés, a entrada de uma camada convolucional também passa pela função de ativação, e só então é repassada para a próxima camada. Muito comumente essa próxima camada é onde é efetuado o *subsampling*. Essas camadas são chamadas de camadas de *pooling*. O *pooling* pode ser feito utilizando diferentes técnicas, como o *average pooling* e o *max pooling*, que estão exemplificadas na Figura 6, mas todas tem o mesmo objetivo, que é criar uma única entrada que sintetize a informação de uma janela de valores do mapa de entrada. O *max pooling*, por exemplo, define como resultado o maior valor de todos os presentes na janela definida. Já o *average pooling* calcula a média da janela e a utiliza como resultado. Independente do método, o objetivo é diminuir a resolução dos mapas de características de modo que somente um sumário da região permaneça, gerando um mapa menor e permitindo uma variação maior da posição exata daquela característica. O processo de

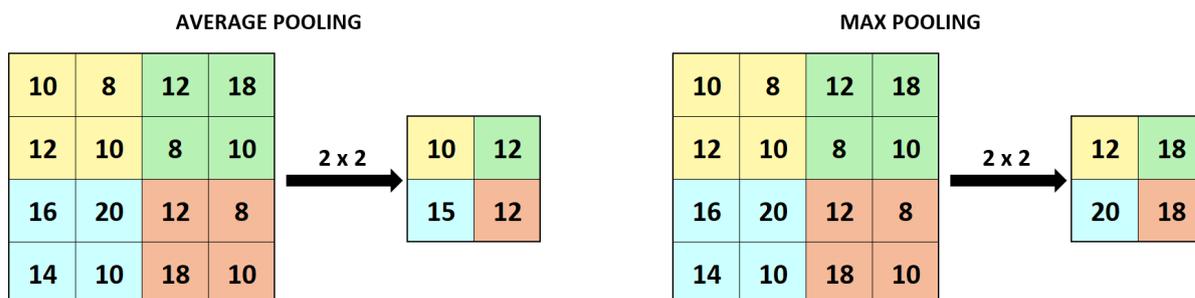


Figura 6 – Comparação entre as técnicas de *Average Pooling* e *Max Pooling* utilizando uma janela de tamanho 2x2 pixels. O método *Average Pooling* calcula a média dos valores da janela, enquanto o método *Max Pooling* seleciona o maior valor discreto do pixel.

*pooling* não é obrigatório e nem sempre é aplicado entre uma camada convolucional e outra.

Os autores em Goodfellow, Bengio e Courville (2016), ainda descrevem que após passar por todas as camadas de convolução, o resultado dos neurônios da última camada convolucional é convertido em uma representação unidimensional. Então, é processado por uma ou mais camadas totalmente conectadas com objetivo uma classificação da entrada em uma das classes pré-definidas, geralmente, tendo sua classificação final definida como uma probabilidade para cada classe, calculada pela função de ativação Softmax.

De modo geral as redes convolucionais são muito importantes pois foram um dos primeiros modelos de *deep learning* a conseguir relevante desempenho, conseguindo superar as RNA para o reconhecimento e classificação de imagens (GOODFELLOW; BENGIO; COURVILLE, 2016).

### 2.1.5 Decomposição de Tensores

O termo tensor é utilizado como uma generalização para variáveis indexadas, independente do número de dimensões. Os tensores com apenas uma dimensão são chamados geralmente de vetores, enquanto os tensores de duas dimensões são comumente chamados de matrizes. De três dimensões em diante, essas estruturas simplesmente são denominadas tensores. O conceito de tensores é bastante utilizado em áreas como processamento de sinais, aprendizado de máquina, mineração de dados, dentre outras (SIDIROPOULOS et al., 2017).

Um tensor pode ter seu formato rearranjado a partir de técnicas de decomposição que conseguem comprimir o número de índices do tensor original. Existem diversas técnicas de decomposição de tensores, sendo duas das mais conhecidas a *Canonical Polyadic Decomposition* (CP Decomposition) e a *Tucker Decomposition*. No geral, essas técnicas consistem em operações algébricas que decompõem o tensor em outros tensores de ordem

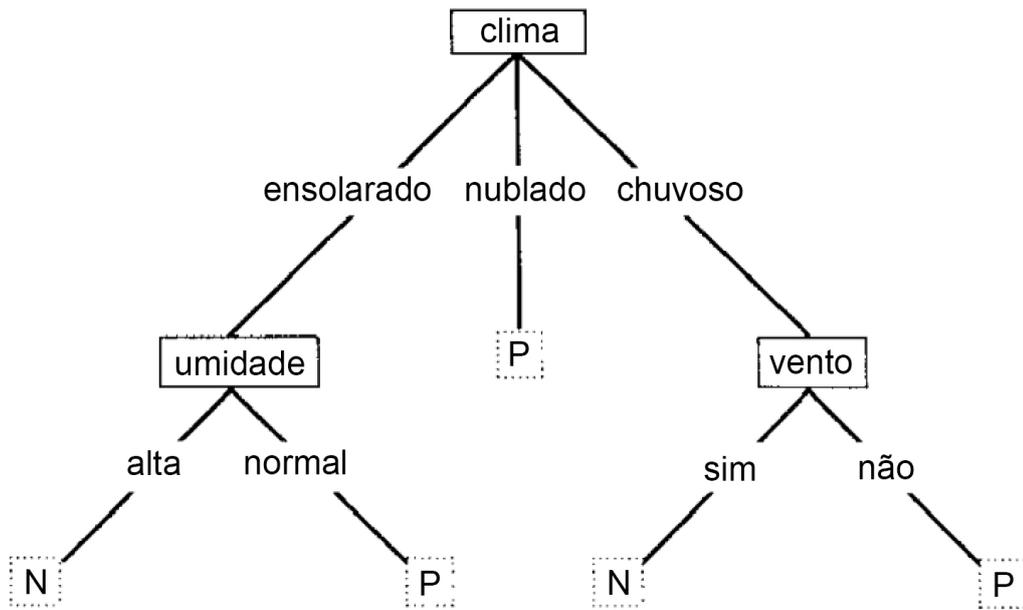


Figura 7 – Uma Árvore de Decisão simples, que classifica, a partir das condições de clima, umidade e vento, um exemplo em classe P (positivo) ou N (negativo). Imagem adaptada de [Quinlan \(1986\)](#).

menor, que podem passar por operações algébricas específicas para a reconstrução que resultam em uma aproximação do tensor que foi decomposto. Esse processo pode facilitar a análise e o armazenamento de grandes conjuntos de dados ([CICHOCKI et al., 2015](#)).

### 2.1.6 Algoritmos de Aprendizado de Máquina

Além da classificação utilizando as camadas totalmente conectadas das arquitetura CNN, também pode-se usar parte das etapas de extração de características para avaliação com os algoritmos de Aprendizado de Máquina (AM). Uma estratégia muito empregada na literatura e obter esses dados nas camadas convolucionais para classificação com os algoritmos clássicos de AM. Os principais algoritmos empregados nesse contexto são: árvores de decisão, florestas aleatórias, máquinas de vetores de suporte e Naive Bayes.

As Árvores de Decisão (do inglês, *decision tree* - DT) foram propostas, inicialmente, por [Quinlan \(1986\)](#) e baseiam-se no conceito de classificar uma amostra a partir de uma estrutura semelhante a uma árvore, composta por nós de decisão e nós terminais. Os nós de decisão levam a distintos possíveis caminhos a partir da análise de uma característica da amostra e os nós terminais efetuam a classificação da amostra em uma das possíveis classes do problema. Possui a vantagem de ser resistente a ruído e ter a capacidade de lidar com exemplos com falta de informações, tanto na etapa de treinamento quanto na etapa de classificação. Uma representação simples de uma Árvore de Decisão pode ser observada na Figura 7.

O algoritmo Floresta Aleatória (do inglês, *random forest* - RF) é método de classificação baseado no uso de múltiplas árvores de decisão que definem um consenso, resultando em apenas uma classificação. Foi proposto por Breiman (2001) e parte do princípio de treinar cada árvore com um diferente subconjunto de amostras do conjunto original. Também são escolhidas de forma aleatória as características que dividem cada nó de cada árvore, garantindo uma maior robustez na classificação final. Para tomar a decisão da classificação de um exemplo, é feito um procedimento baseado em votação, onde a classe mais votada é a eleita como a predição para a amostra sendo classificada.

Uma outra abordagem é a Máquina e Vetores de Suporte (do inglês, *support vector machine* - SVM) a qual baseia-se no conceito de mapear de forma não linear vetores de entrada em um espaço de características, construindo uma decisão não linear a partir deste espaço. Esse algoritmo é amplamente utilizado na classificação de problemas binários, e garante uma boa capacidade de generalização (CORTES; VAPNIK, 1995). Foi inicialmente proposto por Boser, Guyon e Vapnik (1992), porém possuía a capacidade de classificar apenas problemas linearmente separáveis. Posteriormente, Cortes e Vapnik (1995) apresentaram uma nova implementação, capaz de classificar problemas não lineares.

O algoritmo *Naive Bayes* (NB) é um método de aprendizado baseado no Teorema de Bayes que, dado um conjunto de amostras para treino, os descreve de acordo com suas características, e ao fazer uma classificação, baseia-se nas características já aprendidas para classificá-lo na mesma classe das instâncias que mais se aproximam (MITCHELL, 1997). O método é conhecido por sua simplicidade, e por conseguir fazer classificações competentes mesmo partindo de um princípio pouco complexo, alcançando resultados semelhantes ao de outros algoritmos mais elaborados.

### 2.1.7 Métricas de Avaliação de Classificação

Para a avaliação dos resultados obtidos, foi utilizada a acurácia para medir a performance dos modelos. A acurácia é dada pela Equação 2.1, onde: os *true positives* (TP) representam as imagens que foram corretamente identificadas como positivas; *true negative* (TN) representa as imagens corretamente classificadas como negativas; *false positives* (FP) indica as imagens que foram incorretamente indicadas como positivas; e *false negatives* (FN) indica as imagens que foram incorretamente identificadas como negativas.

$$\text{Acurácia} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.1)$$

A medida F é uma métrica que indica a qualidade geral de um método, indepen-

dente do número de classes do conjunto de entrada, ou do desbalanceamento do mesmo. Essa métrica consiste na média harmônica entre duas outras métricas: precisão (*Prec*) e revocação (*Rev*). A *Prec* indica a razão entre amostras classificadas como uma determinada classe A e total de amostras da classe A, dada pela Equação 2.2.

$$Prec = \frac{VP_A}{VP_A + FP_A} \quad (2.2)$$

O cálculo da *Rev* é feito de forma similar, porém os valores de VP e FP utilizados são os referentes às demais classes existentes no problema, conforme indicado na Equação 2.3.

$$Rev = \frac{VP_B}{VP_B + FP_B} \quad (2.3)$$

Desta forma, a medida F, em diversas ocasiões também denominada *F1-Score* é dada pela Equação 2.4.

$$F1 = \frac{2 \times Prec \times Rev}{Prec + Rev} \quad (2.4)$$

## 2.2 Estado da Arte

Nesta seção serão explorados alguns dos trabalhos de outros autores apresentados nos últimos anos. Para fins de organização eles estão separados em duas linhas de pesquisa: i) estudos focados em classificação de displasias; ii) abordagens de classificação que exploram a decomposição de tensores.

### 2.2.1 Classificação de Displasia Auxiliada por Computador

Partindo da premissa de diferenciar dois tipos de lesões na região bucal, o estudo publicado pelos autores em [Banerjee et al. \(2015\)](#), investigou o uso de espectroscopia com infravermelho por transformada de Fourier (FTIR) juntamente com SVM lineares e quadráticas com diferentes combinações de características espectrais para a detecção e diferenciação de leucoplasias orais (OLK) e carcinomas de células escamosas de cavidade oral (OSCC). Os experimentos foram executados em um banco de imagens composto por amostras coletadas de quarenta e sete pacientes diferentes, sendo oito saudáveis, 16 contendo OLK e 23 contendo OSCC. Para o estudo do FTIR foram feitos previamente tratamentos químicos nos tecidos a fim de obter um melhor resultado. Após esses tratamentos, os espectros FTIR foram obtidos com 32 escaneamentos dentro da faixa de 400-4000cm<sup>-1</sup> com uma resolução de 4cm<sup>-1</sup>. Após esse processo, utilizando um método denominado *Forward Feature Selection* (FFS), foram identificadas 18 características mais

relevantes, sendo 6 dessas altamente relevantes para permitir a diferenciação das lesões. Na etapa de classificação, o SVM alcançou o melhor desempenho com a função quadrática, conseguindo classificar das amostras com OLK e OSCC, com uma acurácia média de 89,7%, com sensibilidade de 81,3% e especificidade de 95,7%.

O estudo de [Adel et al. \(2018\)](#) apresenta um método para a detecção e classificação de displasias epiteliais orais utilizando *Oriented FAST* e *Rotated BRIEF* para extração de características e SVM como algoritmo de classificação. O modelo apresentado traz a decisão quanto ao grau da displasia a partir da análise de 16 tipos de características extraídas das imagens microscópicas do tecido epitelial. Algumas dessas características exigem diferentes níveis de magnificação. O método consiste, primeiramente, em uma etapa de pré-processamento, onde as imagens foram modificadas para facilitar a extração de características. Essas modificações consistem em padronizar a resolução de todas as imagens em 600 pixels de altura por 600 pixels de largura e converte-las do padrão RGB para o padrão HSV. Isso ocorre porque o espaço de cores HSV costuma apresentar resultados melhores que o modelo RGB. Após a etapa de pré-processamento, ocorreu a etapa de extração de características em que foram aplicados os métodos *Scale Invariant Feature Transform* (SIFT), *Speeded Up Robust Features* (SURF) e *Oriented FAST and Rotated BRIEF* (ORB). Na etapa de classificação foram testados dois métodos, o SVM e o *K-Nearest Neighbour* (KNN). O *dataset* utilizado nesse trabalho foi gerado pelos próprios autores, consistindo de 46 casos com 3 níveis de magnificação, sendo 7 destes casos saudáveis e os outros 39 com algum grau de displasia. Por fim, a combinação que apresentou melhor resultado foi o modelo proposto utilizando o método ORB para extração de características e o SVM para classificação, proporcionando uma acurácia de 92,6%.

No trabalho apresentado por [Silva et al. \(2022a\)](#) foi explorada uma abordagem que visa a segmentação e quantificação em imagens histológicas de displasia epiteliais na cavidade oral com o uso de redes neurais convolucionais e algoritmos polinomiais. Segundo os autores, esse estudo é possivelmente o primeira investigação, que utiliza métodos computacionais não só para identificar a displasia, mas também para classifica-la em grau de evolução da doença sobre tecidos histológicos. Foram testadas a influência das etapas de pós e pré processamento de imagens na etapa de segmentação. A segmentação foi realizada com uma rede Mask R-CNN, modelo proposto por [He et al. \(2018\)](#). Após a etapa de segmentação e extração dos núcleo, as informações de entropia e morfológicas foram obtidos. Esses descritores foram armazenados em um vetor de característica e foi utilizado como entrada para um algoritmo polinomial para a classificação da imagem. O método proposto por esse trabalho foi avaliado em 296 regiões de interesse (ROI), sendo 74 ROI para cada uma das quatro possíveis categorias: saudável, displasia leve, displasia moderada e displasia severa. O modelo proposto ainda foi investigado sobre diferentes combinações de métodos de pré e pós processamento. O melhor modelo foi o que não utiliza nenhuma técnica de pré processamento e utiliza operações morfológicas após a

segmentação. Com essa abordagem foi possível obter uma acurácia média de 89,31% na etapa de segmentação, ficando a frente de outros métodos como K-mean, SegNet e U-Net. Também mostrou resultados relevantes na etapa de classificação, em que o algoritmo polinomial, alcançou uma acurácia de 92%, sendo superior a outros métodos avaliados.

Buscando uma abordagem diferente para a etapa de extração de características, Rocha et al. (2022) demonstra no trabalho que aplicar o conhecimento de profissionais da saúde para definir o que deve ser utilizado como entrada para uma rede neural artificial simples pode trazer resultados comparáveis a sistemas mais complexos, tendo um custo computacional muito inferior. Portanto, na etapa de extração de características, as regiões a serem apresentadas para a rede de classificação são escolhidas por patologistas, sendo as regiões que possuem as principais características visuais para a classificação da displasia. Na etapa de classificação, foram utilizados diferentes modelos de RNA, todas com uma camada de entrada com 65.536 neurônios, mas utilizando diferentes quantidades de camadas ocultas e diferentes quantidades de neurônios por camada, sendo avaliados modelos com 20, 50, 100 e 150 neurônios por camada ocultas. Como função de custo foram avaliadas tanto a *MSE* quanto a *cross-entropy*. O banco de imagens utilizado para os testes foi obtido do Laboratório de Patologia Oral da Universidade Federal de Santa Catarina e consiste em imagens histológicas da cavidade oral com grau de magnificação de 200 vezes. Ao todo são 172 imagens com resolução de 2584 pixels de largura por 1936 pixels de altura, sendo 88 destas imagens potencialmente de displasias epiteliais, e outras 84 imagens com hiperplasia sem displasia epitelial. O melhor resultado na classificação foi obtido com a rede com três camadas ocultas com 150 neurônios cada e função de ativação *cross-entropy*, alcançando uma média de 87,10% de precisão, valor semelhante ao de uma CNN pré-treinada, que obteve 90,0%, porém com uma especificidade de 93,33% contra 86,67% da pré-treinada. A principal contribuição vem do custo computacional, onde a pré-treinada avaliada, uma ResNet-18, teve um custo computacional de 2 GFlops, enquanto o modelo proposto apenas de 19,79 MFlops.

### 2.2.2 Redes Convolucionais com Decomposição de Tensores

O trabalho publicado por Kim et al. (2016) explora a decomposição de tensores com foco na aplicação em redes convolucionais para torná-las mais leves pensando no uso em dispositivos móveis. O método foi aplicado em diversos tipos diferentes de redes convolucionais composto de três etapas. A primeira consiste na escolha dos valores para os *ranks* da decomposição, o qual foi feita utilizando *variational Bayesian matrix factorization* (VBMF). A segunda etapa, decomposição dos tensores, e a última etapa foi a investigação do *fine-tuning* em toda a rede com o objetivo de recuperar a acurácia perdida durante a decomposição. A técnica de decomposição utilizada foi a Tucker-2 para as camadas convolucionais e a Tucker-1 para as camadas totalmente conectadas. Após

a decomposição é relatado a perda da acurácia, que para ser complementemente recuperada necessitou 10 época ao mínimo na etapa de *fine-tuning*. Os resultados dos experimentos resultaram em redes com 7,40 vezes menor em relação aos parâmetros de pesos e 4,80 vezes menos custo computacional para a rede VGG-S. No caso AlexNet, foi 5,46 vezes menor para os pesos e 2,72 vezes menos custo computacional. A rede AlexNet foi a que mostrou resultados mais promissores. Estes resultados foram obtidos com um impacto mínimo na acurácia.

Buscando estabilidade na aproximação da decomposição, o trabalho apresentado por Phan et al. (2020) propuseram a decomposição de redes convolucionais tanto utilizando *CP decomposition* quanto utilizando uma combinação de *CP decomposition* e *Tucker decomposition*, com a aplicação de uma técnica para a estabilização da decomposição por *CP decomposition* a fim de garantir uma boa recuperação da acurácia da rede na etapa de *fine-tuning*. Na etapa de decomposição das camadas convolucionais com a técnica *CP decomposition*, a decomposição foi aplicada e após foi adotado um método de estabilização. O primeiro método proposto foi uma variante do Error Preserving Correction (EPC), o qual busca preservar o erro de aproximação enquanto ao mesmo tempo busca minimizar a sensibilidade da decomposição. Um segundo método foi proposto para a decomposição das camadas convolucionais com *Tucker decomposition* (TKD) e, em seguida, decompor o núcleo obtido na TKD utilizando a *CP decomposition*. Para camadas especiais com convolução 1x1 pixels, a técnica utilizada foi a Singular Value Decomposition (SVD). Como estimativa para os *ranks* foi utilizado uma heurística baseada em busca binária que seleciona o menor *rank* possível que alcança um limiar de perda mínima de acurácia após o *fine-tuning* de uma única camada. O resultado dos experimentos mostram que o método proposto alcança grandes taxas de compressão nas redes testadas. A VGG-16, por exemplo, testada no *dataset* ILSVRC-12, apresentou uma taxa de compressão de 5,26 vezes menos FLOPS que a rede original, superando os outros métodos avaliados e alcançando uma queda na acurácia.

Com uma abordagem que utiliza *Tensor-Train Decomposition* e estimativa automática de *ranks*, o estudo publicado por Gabor e Zdunek (2022) aplica o método proposto em vários modelos de rede CNN e executa testes em dois banco de imagens diferentes. As redes utilizadas nos experimentos foram TT-conv-CNN, VGGnet, ResNet-56 e a ResNet-110. Os bancos de imagens escolhidos foram o CIFAR-10 e o CIFAR-100. Os *ranks* foram estimados utilizando uma técnica chamada *energy threshold selection*. Na etapa de decomposição esse método divide cada camada convolucional em quatro novas camadas convolucionais, resultantes da decomposição aplicada e a convolução é aproximada utilizando quatro convoluções menores. Todas as camadas convolucionais são decompostas nesse método, exceto a primeira, que os autores julgaram pequena demais. Como resultado, o número de FLOPS e parâmetros das redes foram minimizados, conseguindo taxas de compressão melhores que de outros métodos. Aplicando em uma rede VGGnet, a re-

dução dos parâmetros foi de 3,03 vezes e dos FLOPS de 4,71 vezes, um resultado melhor que com técnicas, como a decomposição CP, que alcançou taxas de 1,93 e 2,00 vezes, ou a DCP, que conseguiu 1,93 e 2,00 vezes para redução de parâmetros e FLOPS, respectivamente. Após a etapa de decomposição foi necessário efetuar uma etapa de *fine-tuning* para que a acurácia pudesse ser recuperada. Na etapa final, a taxa de compressão de FLOPS e parâmetros foram bastante expressiva em detrimento de uma pequena perda de acurácia.

No trabalho de Liu e Ng (2022), foi proposto uma nova técnica para decomposição dos tensores de redes neurais baseado na utilização de decomposição de Tucker com resposta não-linear. O método foi avaliado nas arquiteturas ResNet-18, VGG-16, AlexNet e GoogleNet e foi testado em dois diferentes *datasets*, sendo empregado CIFAR-10 e ILSVRC12. O método consiste em utilizar uma técnica chamada de *Alternating Direction Method of Multipliers* (ADMM), para que seja possível estimar tanto o *rank* para a decomposição quanto a matriz de características da decomposição de Tucker de forma simultânea. Após a aplicação da decomposição, foi feita uma etapa de *fine-tuning*, afim de recuperar a acurácia das redes. Os resultados obtidos foram promissores, apresentando resultados compatíveis ou superiores a trabalhos semelhantes da área. Na arquitetura ResNet-18 avaliando o *dataset* CIFAR-10, a rede teve uma perda na acurácia de apenas 1,56%, enquanto apresentou uma redução no número de parâmetros de 11,82 vezes. Já na arquitetura AlexNet avaliando o *dataset* ILSVRC12, a perda na acurácia foi de apenas 0,37%, alcançando uma taxa de compressão de 7,34 vezes, valor consideravelmente superior ao alcançado por trabalhos da área. Por último, quando aplicada a técnica na arquitetura GoogleNet, a taxa de compressão chega a 1,68 vezes, e utilizando a rede resultante para avaliar o *dataset* ILSVRC12, obtém-se uma acurácia apenas 0,28% menor que a obtida na rede original.

Em um outro estudo proposto por Gabor e Zdunek (2023), uma técnica de decomposição chamada de *Hierarchical Tucker-2* (HT-2) foi aplicada em diversas arquiteturas de redes CNNs e foram executados testes tanto nos bancos de imagens da CIFAR-10 da ImageNet. A decomposição foi efetuada somente nas camadas convolucionais. A compressão foi dividida em duas etapas: (1) As camadas convolucionais, decompostas utilizando o algoritmo de compressão baseado em HT-2; (2) As camadas com *fine-tuning* utilizando um algoritmo de convolução HT-2. Para a escolha dos *ranks* da decomposição, o critério escolhido foi chamado pelos autores de baseado em energia, e consiste no uso de uma fórmula, que envolve valores singulares da matriz a ser decomposta e define os *ranks* a partir da ultrapassagem de um limiar. Para os experimentos, a rede DenseNet-40 foi treinada para avaliar o *dataset* CIFAR-10, e redes pré-treinadas AlexNet, GoogleNet, ResNet-50, ResNet-18 e VGG-16 foram utilizadas para avaliar o *dataset* da ImageNet. Os resultados foram promissores tanto nas redes de média escala quanto nas grandes escalas, conseguindo uma forte redução na taxa FLOPS e parâmetros de cada rede. A Res-Net-50,

por exemplo, ocorreu uma redução de 2,74 vezes em relação aos parâmetros e 2,85 vezes menos FLOPS, sendo esses resultados melhores que todos os outros métodos comparados no estudo. A decomposição, entretanto, causa uma queda na taxa de acurácia da rede, mas pode ser recuperada através de uma etapa de *fine-tuning*.

### 2.2.3 Considerações Finais

Os trabalhos apresentados demonstram que o uso de decomposição de tensores para a compressão de redes neurais tem apresentado resultados promissores, alcançando taxas de compressão significativas, ao mesmo tempo que mantém a acurácia muito próxima da original. Porém, dos trabalhos citados que se propõe a classificar displasia, nenhum avalia técnicas de compressão de rede. Da mesma forma, alguns dos trabalhos citados que avaliam técnicas para compressão de redes utilizam técnica de decomposição como a decomposição de Tucker, como os trabalhos apresentados por [Kim et al. \(2016\)](#) e [Liu e Ng \(2022\)](#). No entanto, nenhum avalia as redes resultantes em *datasets* de imagens histológicas. Em razão disso, este trabalho propõe avaliar os efeitos da aplicação da decomposição de Tucker em redes neurais destinadas a classificação de imagens histológicas.

## 3 Materiais e Métodos

Nesta seção são apresentados os conceitos relacionados os métodos utilizados no estudo, como as características dos bancos de imagens. Também há detalhes dos processos das redes neurais convolucionais relacionados ao treinamento e compressão, assim como as *features* extraídas e utilizadas em algoritmos de ML e métricas de avaliação. Na Figura 8 pode-se observar um diagrama exemplificando as etapas aplicadas para o desenvolvimento da ferramenta proposta. Todos os testes conduzidos neste trabalho foram executados em uma máquina com processador AMD Ryzen 5 3600XT, placa de vídeo GeForce RTX 2070 SUPER e 64Gbs de memória RAM e utilizaram linguagem Python.

### 3.1 Banco de imagens

Este trabalho apresenta uma investigação de algoritmos computacionais para imagens histológicas de displasia. Nesses experimentos foram empregados dois diferentes bancos de imagens histológicas extraídos de lâminas de tecidos da cavidade oral.

O primeiro banco de imagens utilizado neste trabalho, referido como banco de imagens Silva, de domínio privado, é composto de imagens histológicas de diferentes graus de displasia, obtidas a partir de línguas de camundongos C57BL/6, que foram lesionadas experimentalmente utilizando o carcinógeno 4-nitroquinolina-N-óxido (4NQO), diluído na água dos animais. Esses experimentos foram devidamente aprovados pelo Comitê de Ética na Utilização de Animais sob o número 038/09 na Universidade Federal de Uberlândia. Foram utilizados um total de 30 camundongos, que foram divididos em dois grupos de mesmo tamanho, um para abate após 16 semanas a partir do início do experimento, e

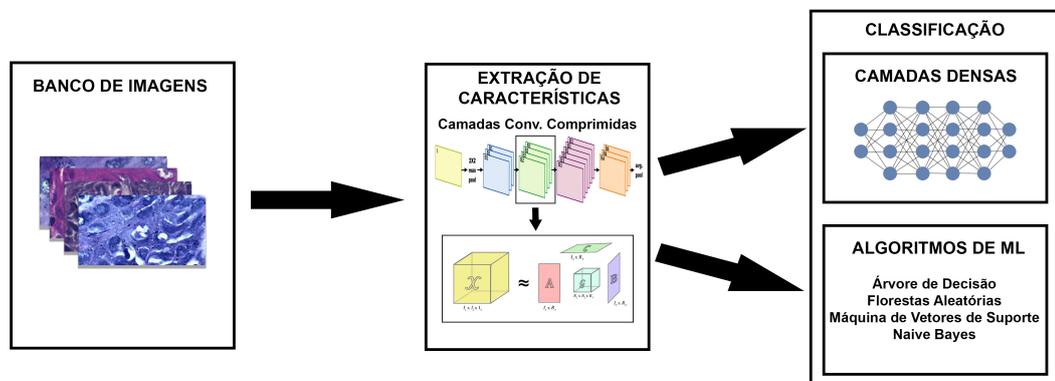


Figura 8 – Etapas dos métodos e materiais empregados para classificação das lesões nas imagens histológicas.

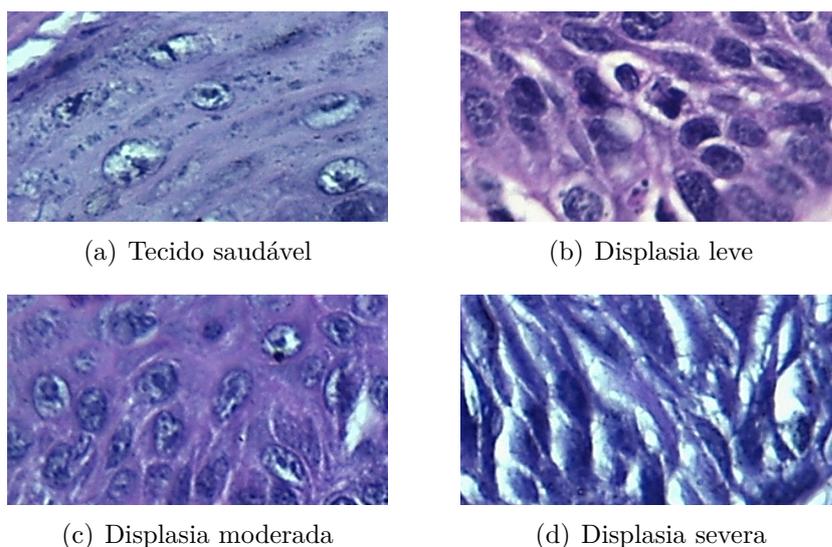


Figura 9 – Exemplos de imagens do banco elaborado por Silva.

outro para abate após 32 semanas. Durante as 16 primeiras semanas todos os animais beberam a água com o carcinógeno diluído e na 16<sup>a</sup> semana os espécimes do primeiro grupo foram sacrificados. Nas próximas semanas os animais restantes beberam somente água filtrada e foram sacrificados na 32<sup>a</sup> semana (SILVA, 2019). Após o abate, os camundongos tiveram suas línguas removidas e devidamente tratadas para a composição das lâminas para classificação. Esses experimentos foram realizados na Área de Patologia Oral, da Faculdade de Odontologia, da Universidade Federal de Uberlândia. As lâminas com tecidos corados foram digitalizadas a partir de um microscópio ótico modelo Leica DM500, utilizando um grau de magnificação de 400x. Essas imagens foram classificadas por dois especialistas entre tecido saudável, displasia leve, displasia moderada e displasia severa, conforme o modelo definido por Lumerman, Freedman e Kerpel (1995). Um total de 66 imagens com resolução de 2048 x 1536 pixels foram obtidas, a partir das quais obteve-se 74 ROI's com resolução de 450 x 250 pixels para cada uma das classes nas regiões com presença da lesão. No caso de tecidos saudáveis, essas informações foram extraídas no centro da imagem. Um exemplo de cada classe do banco pode ser observado na figura 9.

O segundo banco de imagens utilizado, denominado banco de imagens Costa, de domínio privado, foi construído a partir de lesões obtidas experimentalmente com a exposição ao carcinógeno 4NQO, diluído na água para os roedores da cepa B6.129P2-Nos2. Todos os procedimentos executados foram devidamente aprovados pela Comitê de Ética na Utilização de Animais sob o número 100/18. O tratamento durou 16 semanas, e os animais ficaram em observação por mais 8 semanas, sendo eutanasiados após um período de 24 semanas desde o início do experimento. Os experimentos também foram realizados na Área de Patologia Oral, da Faculdade de Odontologia, da Universidade Federal de Uberlândia.

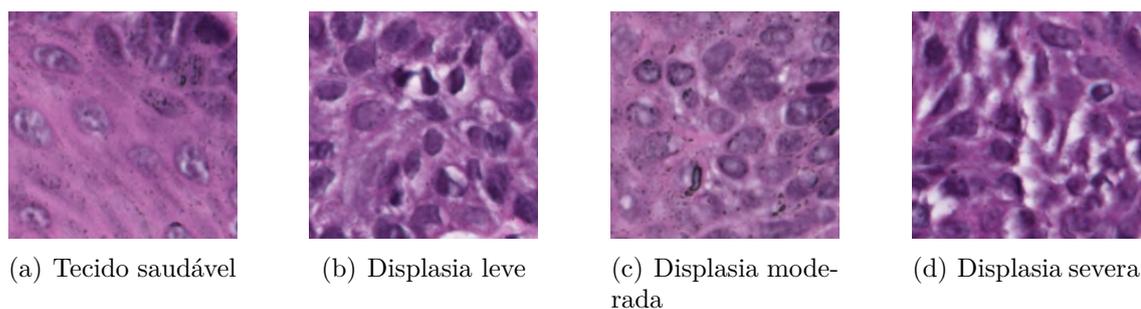


Figura 10 – Exemplos de imagens do banco elaborado por Costa.

As classificações foram feitas por dois patologistas, a partir da análise na língua dos roedores, e foram categorizadas nos seguintes níveis de alteração: displasia leve, displasia moderada, displasia severa e tecido saudável. As lâminas foram digitalizadas no equipamento Leica Aperio AT2 com 400x de magnificação. As regiões de interesse (ROIs) foram extraídas utilizando o software QuPath e com resolução de 200 x 200 pixels nas regiões com presença das lesões. Em tecidos saudáveis, as ROIs foram extraídas no centro da imagem. Foram obtidas 215 ROIs classificadas como tecido saudável, 93 como displasia leve, 141 como displasia moderada e 110 como displasia severa, somando um total de 559 imagens digitais. A Figura 10 mostra um exemplo de cada classe do Banco Costa.

A Tabela 1 apresenta uma relação da quantidade de imagens de cada classe em cada um dos bancos de imagem.

Tabela 1 – Número de imagens de cada classe presentes em cada banco de imagem.

	Tecido Saudável	Displasia Leve	Displasia Moderada	Displasia Severa	Total
Banco Silva	74	74	74	74	296
Banco Costa	215	93	141	110	559

## 3.2 Aumento de Dados

Devido ao número reduzido de amostras em cada um dos bancos de imagens, foram aplicadas técnicas de aumento de dados (do inglês, *data augmentation* - DA) para garantir uma melhor generalização das redes treinadas. Foram aplicados as operações de espelhamento vertical e horizontal, com uma chance de 50% por imagem, e rotação aleatória, variando de um ângulo de 0° a 40°. Além disso, foram aplicados as operações de recortes aleatórios na imagem, que variam de 0.80 a 0.90 do tamanho da imagem, com operação após o corte de um *resize* para a resolução original das imagens. Também foi aplicado auto contraste e *sharpness* em todas as imagens a fim garantir um melhor contraste entre os elementos das imagens.

Por último, foram aplicadas modificações aleatórias no brilho, contraste e saturação das imagens, variando entre 0.70 e 1.30 dos valores originais. As mudanças de contraste também foram empregadas para garantir que a rede conseguisse uma boa generalização, mesmo com as variações de cor que ocorrem entre as imagens. Na Figura 11 pode-se observar exemplos destas transformações aplicadas na sobre a mesma amostra, onde todas as transformações foram aplicadas, com valores definidos aleatoriamente, dentro dos intervalos definidos.

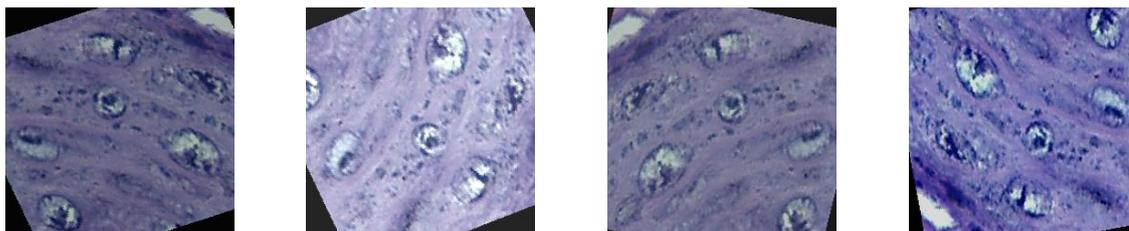


Figura 11 – Exemplos de transformações para aumento de dados.

### 3.3 Arquitetura ResNet

As *Residual Networks* (ResNet) foram propostas, inicialmente, por (HE et al., 2015) e se baseiam no conceito chamado de aprendizado residual, que consiste no uso de saltos de uma ou mais camadas, mantendo a informação intacta e reinserindo-a no final de camadas subsequentes. O uso dessa abordagem procura minimizar o problema de degradação que ocorre em redes muito profundas, garantindo uma performance superior quando comparada a outras arquiteturas que não utilizam esta técnica. A Figura 12 ilustra a arquitetura da ResNet-50.

A escolha dessa arquitetura deve-se as motivações: i) amplamente utilizada em outros trabalhos de imagens histológicas, como para a classificação de diferentes partes do corpo (TALO, 2019), para a classificação de imagens de câncer colo-retal (SARWINDA et al., 2021), e para a classificação de imagens de câncer de mama (AL-HAIJA; ADEBANJO, 2020); ii) a sua estrutura, que é composta majoritariamente por camadas convolucionais. Como a decomposição de Tucker é aplicada apenas nas camadas convolucionais, utilizar uma rede que possui relativamente poucos parâmetros em suas camadas densamente conectadas, ajuda a garantir que a compressão seja mais expressiva.

A arquitetura ResNet está disponível em diferentes versões, que possuem diferentes números de camadas. Para a execução dos experimentos propostos neste trabalho foram escolhidos os modelos ResNet-18 e ResNet-50, que possuem, respectivamente, 18 e 50 camadas.

Tanto a arquitetura ResNet-18 quanto a ResNet-50, independente de qual conjunto de dados utilizado, foram treinadas por 500 épocas. O otimizador escolhido foi o *Stochastic*

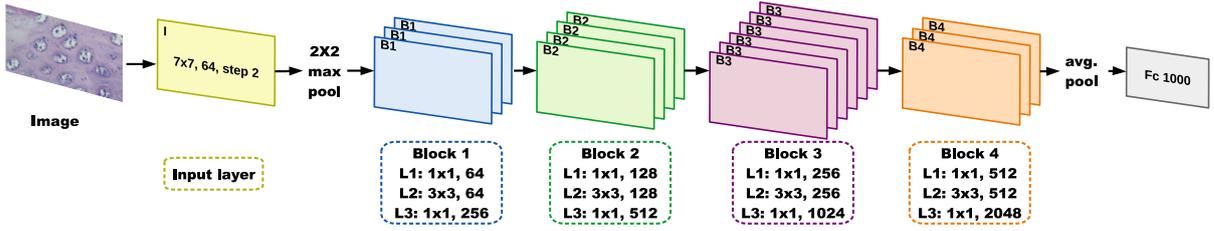


Figura 12 – Arquitetura de uma ResNet-50. Imagem extraída de [Silva et al. \(2022a\)](#).

*Gradient Descent* (SGD) e a função de perda empregada foi a de Entropia Cruzada. O *learning rate* (lr) utilizada em ambas as arquiteturas na etapa de treinamento foi de 0.001.

### 3.4 Decomposição de Tucker

Como uma extensão da decomposição de valores singulares (do inglês, *singular value decomposition* - SVD) para variáveis indexadas de ordem maior do que duas dimensões, a decomposição de Tucker funciona basicamente executando operação nos espaços ortogonais que estão associados as diferentes  $n$  dimensões de um tensor, gerando  $n$  matrizes, além de um núcleo com as dimensões dos *ranks* escolhidos, que quando manipulados podem resultar em uma reconstrução que se aproxima do tensor original. Quando essa decomposição resulta em um número total de parâmetros menor, que o número do tensor original, é denominada *low-rank approximation* ([KIM et al., 2016](#)).

Para tanto, define-se o *rank* de um tensor  $\mathcal{X}$ , o menor valor possível dado por *rank-1*, que quando somados resultam em exatamente  $\mathcal{X}$  decomposto. Então, o *rank* é o menor número de componentes necessários para uma decomposição que gera uma reconstrução ao tensor original. Quando uma decomposição utiliza como dado de *rank*, o valor de *rank* do tensor a ser decomposto, pode-se chamar de *rank decomposition* ([KOLDA; BADER, 2009](#)). No entanto, utilizar o valor de *rank* na decomposição para uma reconstrução exata pode ser custoso computacionalmente, consumindo muita memória e pode resultar em uma decomposição que não seja significativamente mais compacta. Por isso, pode-se utilizar um valor de *rank* menor que o *rank* do tensor a ser decomposto, resultando em uma decomposição que não será reconstruída, mas não sofrerá degradações conforme descrito por ([KIM et al., 2016](#)).

A Equação 3.1 define a decomposição de Tucker para um tensor  $\mathcal{X}$  de três dimensões, cujas dimensões tem valores  $I_1$ ,  $I_2$  e  $I_3$ . Portanto,  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  e os valores escolhidos para *rank* são  $R_1$ ,  $R_2$  e  $R_3$ :

$$\mathcal{X} \approx \mathcal{G} \times_1 A \times_2 B \times_3 C. \quad (3.1)$$

Após a decomposição, obtém-se três matrizes,  $A \in \mathbb{R}^{I_1 \times R_1}$ ,  $B \in \mathbb{R}^{I_2 \times R_2}$ ,  $C \in \mathbb{R}^{I_3 \times R_3}$ , e um núcleo tridimensional  $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ . Essas matrizes podem ser observadas

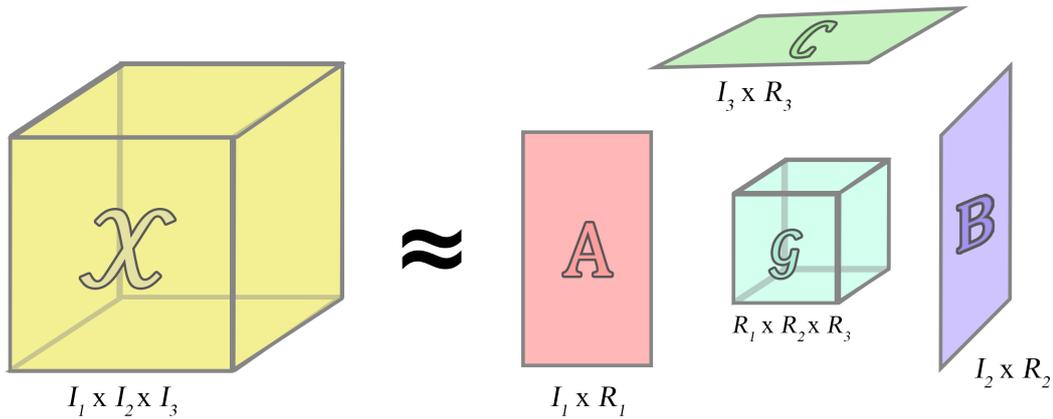


Figura 13 – A decomposição de Tucker aplicada em um tensor de três dimensões, resultando em três matrizes e um núcleo.

de maneira gráfica na Figura 13, a qual ilustra o tensor original e o resultado da sua decomposição. Utilizando as matrizes  $A$ ,  $B$  e  $C$ , e o núcleo  $G$ , é possível então reconstruir uma aproximação do tensor original.

Então, após o treino das redes neurais convolucionais, a decomposição de Tucker foi aplicada as arquiteturas. As operações de decomposição foram aplicadas somente nas camadas convolucionais das redes. Como os tensores de cada camada possuem tamanhos variados, o parâmetro para escolha do *rank* de cada tensor foi definido de forma proporcional, sendo um número decimal entre 0 e 1, que é multiplicado pelo valor original das dimensões do tensor, resultando no valor de *rank* a ser aplicado na decomposição. Caso o valor obtido para *rank* seja não seja um número inteiro, o valor é arredondado. Para os experimentos executados, os valores de proporção para os *ranks* foram definidos como 0,50, 0,40, 0,30, 0,20 e 0,10 em relação aos valores das dimensões originais. Esse valores foram utilizados para a ResNet-50 e ResNet-18.

### 3.5 Avaliação de Desempenho das Abordagens

Para a execução dos experimentos, os bancos de imagens foram avaliados sempre de maneira binária, sendo divididos em duas abordagens diferentes: a primeira consistiu em utilizar apenas as classes de tecido saudável e displasia severa, mantendo o *dataset* com um número balanceado de imagens. A segunda abordagem, consistiu em definir uma classe como saudável e outra como displasia, contendo as imagens de displasia leve, moderada e severa, formando um conjunto de imagens desbalanceado em que há um número maior de amostras com a lesão. Desta forma, os modelos CNNs foram avaliados em quatro conjuntos de imagens denominados nesse trabalho: Conjunto Silva Balanceado, contendo 148 imagens (74 saudável e 74 displasia severa); Conjunto Silva Desbalanceado (74 sau-

Tabela 2 – Parâmetros dos diferentes algoritmos classificadores implementados em scikit-learn

Algoritmos	Parâmetros
DT	criterion = "gini", splitter = "best", max_depth = None, min_samples_split = 2
RF	n_estimators = 100, criterion = "gini", max_depth = None, min_samples_split = 2
SVM	C = 1.0, kernel = "rbf", degree = 3, gamma = 'scale'
NB	prior = None, var_smoothing = 1e-9

dável e 222 displasias), contendo 296 imagens; Conjunto Costa Balanceado, contendo 220 imagens (110 saudável e 110 displasia severa); Conjunto Costa Desbalanceado, contendo 559 imagens (215 saudável e 334 displasias).

Após o treinamento das redes originais, a decomposição de Tucker foi empregada. As operações de decomposição foram aplicadas somente às camadas convolucionais. Após a etapa de de composição foi feita uma etapa de *fine-tuning* na rede, para garantir a recuperação da acurácia.

Além de avaliar o desempenho das redes utilizando as camadas densamente conectadas, os bancos de imagens também foram classificados utilizando os algoritmos clássicos de ML. Para isso, as *features* foram extraídas das redes convolucionais após sua vetorização na camada de *average pooling*, sem aplicação das camadas densas e a função Softmax. As *features* foram utilizadas como entrada para treinar e avaliar os algoritmos ML. Essa abordagem foi utilizada para avaliar como a decomposição dos tensores das camadas convolucionais influencia o desempenho em modelos tradicionais em relação as camadas densamente conectadas. Os modelos escolhidos das CNNs investigadas foram baseadas nos resultados obtidos após aplicação das decomposições Turcker dos tensores. Na Tabela 2 estão descritos os parâmetros utilizados em cada um dos algoritmos avaliados.

Nestes experimentos foram empregados a métrica acurácia para avaliação da classificação dos modelos. Também foi utilizado como métrica de avaliação o número total de parâmetros, o número de parâmetros nas camadas convolucionais, a taxa de compressão da rede, tempo de execução com e sem a execução do *fine-tuning*.

## 4 Resultados

Os modelos de CNN foram avaliados para a classificação de imagens histológicas da cavidade oral utilizando os diferentes bancos de imagens. As *features* extraídas das última camada antes das camadas densamente conectadas das CNNs também foram avaliadas em conjunto com algoritmos de ML.

### 4.1 Classificação com os modelos ResNet com Banco de Imagens Silva

As Tabelas 3 e 4 apresentam os resultados obtidos no conjunto de imagens Silva Balanceado e conjunto de imagens Silva Desbalanceado, respectivamente. Para o modelo ResNet-18 (Tabela 3), o número total de parâmetros era cerca de 11 milhões. Após a etapa de decomposição, utilizando o valor de proporção de *rank* de 0,50, o número de parâmetros diminuiu para aproximadamente 4 milhões, resultando em uma redução de cerca de 2,66 vezes. Com o *rank* como 0,40 da proporção original, o número de parâmetros foi para aproximadamente 3 milhões, o que equivale a uma redução de cerca de 3,75 vezes. Na

Tabela 3 – Valores obtidos com as redes convolucionais decompostas avaliando o banco de imagens Silva balanceado.

Banco de Imagens Silva Balanceado						
Rank	Acurácia (%)	F1-Score (%)	Parâmetros	Taxa de Compressão	Tempo para Decompor (segundos)	Tempo para Fine-tune (segundos)
ResNet-18						
Original	100	100	11.177.538	1	-	-
0,50	100	100	4.205.890	2,66	7,24	57,11
0,40	85,71	85,41	2.978.663	3,75	7,01	46,01
0,30	92,85	92,82	1.953.645	5,72	6,86	43,65
0,20	92,85	92,82	1.152.867	9,70	5,98	41,91
0,10	100	100	566.482	19,73	3,87	40,45
ResNet-50						
Original	100	100	23.512.130	1	-	-
0,50	100	100	28.566.722	0,82	13,08	116,71
0,40	100	100	22.638.820	1,04	12,41	108,04
0,30	92,85	92,82	17.078.229	1,38	11,66	98,55
0,20	100	100	11.933.648	1,97	10,37	87,22
0,10	100	100	7.200.064	3,27	6,34	76,70

Tabela 4 – Valores obtidos com as redes convolucionais decompostas avaliando o banco de imagens Silva desbalanceado.

Banco de Imagens Silva Desbalanceado						
Rank	Acurácia (%)	F1-Score (%)	Parâmetros	Taxa de Compressão	Tempo para Decompor (segundos)	Tempo para Fine-tune (segundos)
ResNet-18						
Original	96,42	94,99	11,177,538	1	-	-
0,50	96,42	94,99	4.205.890	2,66	7,12	102,48
0,40	92,85	90,47	2.978.663	3,75	7,15	91,88
0,30	96,42	94,99	1.953.645	5,72	6,95	87,35
0,20	89,28	86,34	1.152.867	9,70	5,84	83,01
0,10	92,85	90,47	566.482	19,73	3,41	80,25
ResNet-50						
Original	96,42	95,44	23.512.130	1	-	-
0,50	92,85	90,47	28.566.722	0,82	12,92	230,90
0,40	89,28	86,34	22.638.820	1,04	12,42	208,15
0,30	92,85	89,33	17.078.229	1,38	11,81	191,66
0,20	92,85	89,33	11.933.648	1,97	10,69	171,77
0,10	89,28	85,71	7.200.064	3,27	6,70	147,17

decomposição utilizando o valor de *rank* de 0,30, a rede resultando apresentou menos de 2 milhões de parâmetros, representando uma compressão de 5,72 vezes. Utilizando o valor de *rank* de 0,20, a redução de parâmetros chegou a marca de 1,5 milhões, equivalendo a uma compressão de 9,70 vezes. Com o valor de *rank* de 0,10, o modelo resultou possui apenas 500 mil parâmetros, o que equivale a uma compressão de mais de 19 vezes. Em relação à acurácia da a rede, a rede com *rank* 0,50 manteve os 100% de acurácia, enquanto o modelo com *rank* 0,40 alcançou 85,71% e os modelos com 0,30 e 0,20 obtiveram acurácia de 92,85% e 92,82 de F1-score. Por último, o modelo resultante da decomposição com *rank* 0,10 manteve a acurácia de 100%.

Em relação ao modelo ResNet50 (ver Tabela 3), o número inicial de parâmetros era de cerca de 23,5 milhões. Após ser decomposta com o valor de *rank* de 0,50 em relação ao original, o modelo resultante sofreu um aumento na quantidade de parâmetros, resultando em 28,5 milhões. Utilizando como *rank* uma proporção de 0,40, o número de parâmetros foi para cerca de 22,6 milhões, apresentando uma pequena diminuição em relação ao modelo original. No modelo resultante da decomposição com *rank* de 0,30, o número de parâmetros reduziu para aproximadamente 17 milhões, significando uma proporção de 1,38 vezes em relação a rede original. Utilizando proporção de *rank* de 0,20, a rede resultante obteve pouco menos de 12 milhões de parâmetros, o que equivale a uma compressão de 1,98 vezes. Por último, o modelo resultante da decomposição com proporção de *rank* de 0,10 apresentou um número total de cerca de 7 milhões de parâmetros,

Tabela 5 – Valores obtidos com as redes convolucionais decompostas avaliando o banco de imagens Costa balanceado.

Banco de Imagens Costa Balanceado						
Rank	Acurácia (%)	F1-Score (%)	Parâmetros	Taxa de Compressão	Tempo para Decompor (segundos)	Tempo para Fine-tune (segundos)
ResNet-18						
Original	100	100	11.177.538	1	-	-
0,50	95,45	96,45	4.205.890	2,66	7,26	118,05
0,40	100	100	2.978.663	3,75	7,26	100,88
0,30	100	100	1.953.645	5,72	6,77	93,26
0,20	95,45	96,45	1.152.867	9,70	5,74	90,48
0,10	100	100	566.482	19,73	3,52	88,34
ResNet-50						
Original	100	100	23.512.130	1	-	-
0,50	100	100	28.566.722	0,82	13,24	251,72
0,40	100	100	22.638.820	1,04	12,53	229,30
0,30	100	100	17.078.229	1,38	11,61	211,84
0,20	100	100	11.933.648	1,97	10,29	187,89
0,10	100	100	7.200.064	3,27	6,43	162,99

significando uma compressão de aproximadamente 3,29 vezes. Em relação à acurácia, todas as redes conseguiram manter o valor em 100%, exceto a resultante da decomposição com proporção de 0,30, que obteve 92,82%.

No caso do banco de imagens Silva desbalanceada (ver Tabela 4), o modelo ResNet-18 originalmente alcançava uma acurácia de 96,42%. Ao ser decomposto, o modelo resultante da decomposição com *rank* de proporção 0,20 obteve uma acurácia de 89,28%, enquanto os modelos decompostos com *rank* 0,40 e 0,10 obtiveram uma acurácia de 92,85%. Os modelos que alcançaram a maior acurácia foram o resultantes da decomposição com *rank* proporcionais a 0,50 e 0,30 em relação as dimensões originais, que obtiveram uma acurácia de 96,42%, mesma acurácia do modelo original. Em relação ao modelo ResNet-50, a rede original alcançava uma acurácia de 96,42%. Após a decomposição com os valores de *rank* de 0,50, 0,30 e 0,20, a acurácia alcançada foi de 92,85%. Quando decomposta com valor de *rank* de 0,40 e 0,10, a acurácia obtida foi de 89,28%. Nenhuma das redes resultantes das decomposições conseguiu manter a acurácia da rede original.

## 4.2 Resultados no Banco de Imagens Costa com a ResNet

As Tabelas 5 e 6 apresentam os resultados obtidos no conjunto de imagens Costa Balanceado e conjunto de imagens Costa Desbalanceado, respectivamente.

Tabela 6 – Valores obtidos com as redes convolucionais decompostas avaliando o banco de imagens Costa desbalanceado.

Banco de Imagens Costa Desbalanceado						
Rank	Acurácia (%)	F1-Score (%)	Parâmetros	Taxa de Compressão	Tempo para Decompor (segundos)	Tempo para Fine-tune (segundos)
ResNet-18						
Original	96,36	96,21	11.177.538	1	-	-
0,50	92,72	92,52	4.205.890	2,66	7,13	188,15
0,40	98,18	98,09	2.978.663	3,75	7,01	169,87
0,30	98,18	98,09	1.953.645	5,72	6,94	160,69
0,20	94,54	94,35	1.152.867	9,70	5,89	154,75
0,10	98,18	98,09	566.482	19,73	3,48	151,44
ResNet-50						
Original	96,36	96,21	23.512.130	1	-	-
0,50	92,72	92,52	28.566.722	0,82	13,20	427,03
0,40	90,90	90,71	22.638.820	1,04	12,47	385,28
0,30	89,09	88,91	17.078.229	1,38	11,69	360,44
0,20	90,90	90,59	11.933.648	1,97	10,51	322,32
0,10	89,09	88,91	7.200.064	3,27	6,52	282,52

No Banco Costa Balanceado (Tabela 5), a ResNet-18 original alcançava uma acurácia de 100% e 100 de F1. Após a decomposição, todos os modelos conseguiram manter a acurácia de 100%, exceto o modelo decomposto com a proporção de *rank* de 0,20, que obteve uma acurácia de 96,36%. Em relação a arquitetura ResNet-50, a rede original tinha uma acurácia de 100%. Após o processo de compressão, apenas a rede resultante da decomposição com proporção de *rank* 0,50 obteve queda na acurácia, resultando em 96,36%. Todas as outras redes resultados conseguiram manter os 100% de acurácia.

A acurácia do modelo ResNet-18 original no banco de imagens Costa Desbalanceado era de 96,36% (Tabela 6). Os modelos que foram decompostos utilizando proporções de *rank* de 0,40, 0,30 e 0,10 conseguiram melhorar a acurácia para um valor de 98,18%. A acurácia nos modelos com *rank* 0,50 e 0,20 foi, respectivamente, de 92,72% e 94,54%. No modelo ResNet-50, a rede originalmente alcançava uma acurácia de 96,36%. O modelo resultante da decomposição com *rank* 0,50 obteve-se 92,72% de acurácia. Já os modelos resultantes da decomposição com os valores de *rank* de 0,40 e 0,20 alcançaram 90,90% de acurácia. Por último, os modelos que utilizaram *rank* 0,30 e 0,10 alcançaram uma acurácia de 89,09% e F1-Score de 88,91.

### 4.3 Classificação com os Algoritmos de ML

As redes escolhidas para a extração das *features* foram aquelas que obtiveram os melhor resultado para cada arquitetura e banco de imagens. Para o banco Silva Balanceado, foram escolhidos os modelos obtidos com a decomposição utilizando *rank* 0,10, tanto para a ResNet-18 quanto para a ResNet-50. Para o banco Silva Desbalanceado, os modelos escolhidos foram os decompostos com *rank* 0,30 e 0,20, respectivamente, para a ResNet-18 e ResNet-50. No caso do banco Costa Balanceado foi escolhida a decomposição de *rank* 0.10 para as duas arquiteturas. Para o Banco Costa Desbalanceado foi escolhido a rede decomposta com *rank* 0,10 para ResNet-18 e com *rank* 0,20 para a ResNet-50. Os algoritmos de ML escolhidos para serem avaliados foram DT, RF, SVM e NB. As implementações foram feitas em linguagem Python, utilizando a biblioteca *scikit-learn*.

Na Figura 14 é mostrado os resultados obtidos utilizando as *features* extraídas das redes treinadas no banco Silva Balanceado. É possível observar que no modelo ResNet-18 houve uma melhora no desempenho do algoritmo DT após a decomposição, aumentando a acurácia de 92,9% para 100%. Para os outros algoritmos, ambas as redes conseguiram acurácia de 100%. Em relação ao modelo ResNet-50, houve uma melhor de desempenho no algoritmo NB após a decomposição, que antes atingia uma acurácia de 85,7% e depois obteve 100%. Todos os outros algoritmos conseguiram 100% de acurácia em ambas as arquiteturas ResNet.

Na Figura 15 é apresentado os resultados obtidos pelos algoritmos de ML que utilizaram as *features* extraídas das redes treinadas no banco Silva Desbalanceado. Em relação a ResNet-18, Utilizando as *features* da rede original os algoritmos DT, RF e SVM apresentaram uma acurácia de 96,4%. Ao serem utilizados em conjunto com a rede comprimida, a acurácia alcançada foi de 92,9%. Já o algoritmo NB manteve a mesma acurácia com ambas as redes, que foi de 92,9%. Em relação ao modelo ResNet-50 original, os algoritmos DT, RF SVM e NB alcançaram, respectivamente, 96,4%, 100%, 100% e 89,3%. Com o ResNet-50 comprimidos os valores obtido foram de 92,9%, 100%, 96,4% e 82,1%, sendo o algoritmo RF o único a manter o valor da acurácia.

É possível observar na Figura 16 os resultados obtidos pelos algoritmos de ML utilizando as *features* extraídas da redes treinadas com o banco Costa Balanceado. Com o modelo ResNet-18, tanto com a rede original quanto com a rede comprimida, os algoritmos alcançaram os mesmo resultados, que foi de 100% de acurácia em todos os algoritmos testados. Em relação a arquitetura ResNet-50, ao utilizar as *features* da rede original, todos os algoritmos também obtiveram 100% de acurácia. Ao serem utilizados em conjunto com a rede comprimida, o resultado se manteve.

Na Figura 17 é apresentado os resultados obtidos pelos algoritmos ao serem treinados utilizados as *features* extraídas das redes treinadas no banco Costa Desbalanceado.



(a) Modelos ResNet-18 Original e Comprimido.



(b) Modelos ResNet-50 Original e Comprimido.

Figura 14 – Comparação entre os resultados obtidos nos algoritmos de ML avaliando o Banco Silva Balanceado.

Com o modelo ResNet-18 original os algoritmos DT, RF, SVM e NB alcançaram, respectivamente, 98,2%, 96,4%, 96,4% e 96,4%. Ao utilizar as *features* da rede comprimida, as acurácias obtidas foram de 94,5%, 96,4%, 92,7% e 90,9%, sendo o algoritmo RF o único que manteve a acurácia original. Ao utilizar as *features* extraídas do modelo ResNet-50 original, os resultados obtidos foram 90,9%, 96,4%, 98,2% e 87,3%, respectivamente para os algoritmos DT, RF, SVM e NB. Ao utilizar as *features* do modelo comprimido, os resultados obtidos foram 83,6%, 92,7%, 90,9% e 87,3%, sendo o NB o único algoritmo a manter a mesma acurácia.

## 4.4 Análise de Desempenho em Relação aos Estudos

A Tabela 7 contém uma análise entre diferentes métodos propostos para a classificação de imagens histológicas de displasia na cavidade oral. Os resultados mostram que a abordagem investigada contribui na classificação de imagens histológicas na cavidade

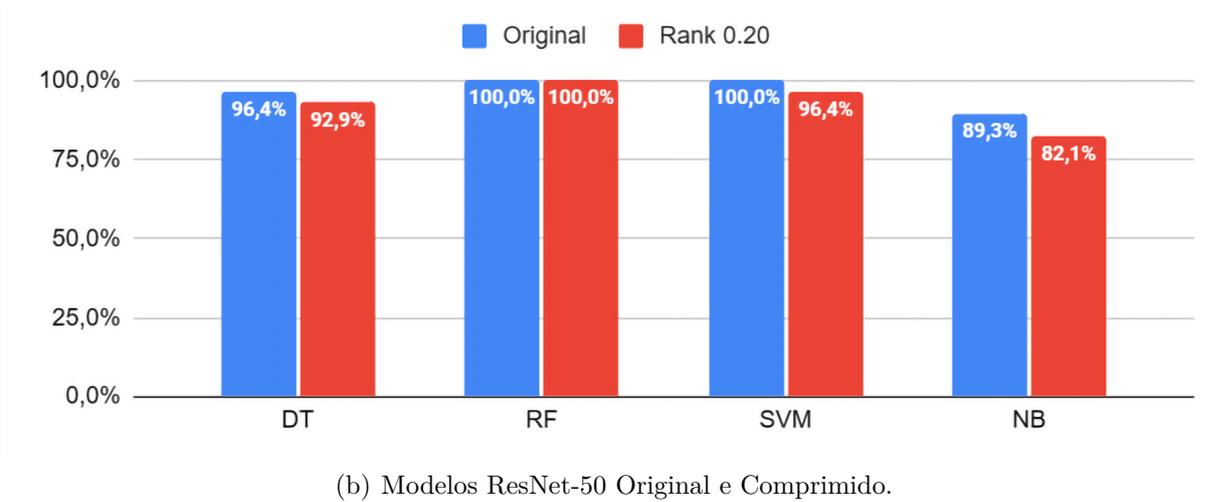
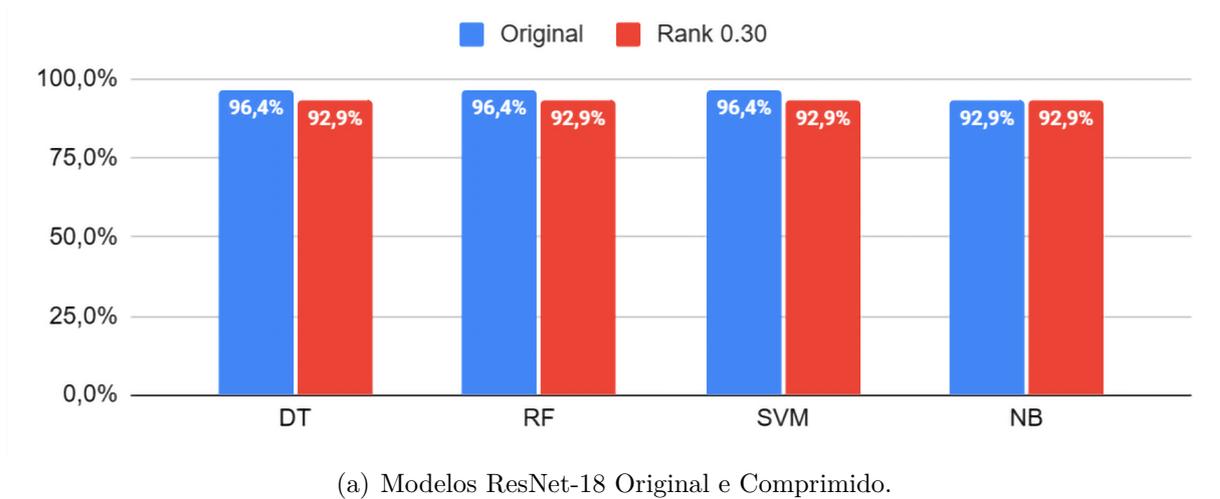
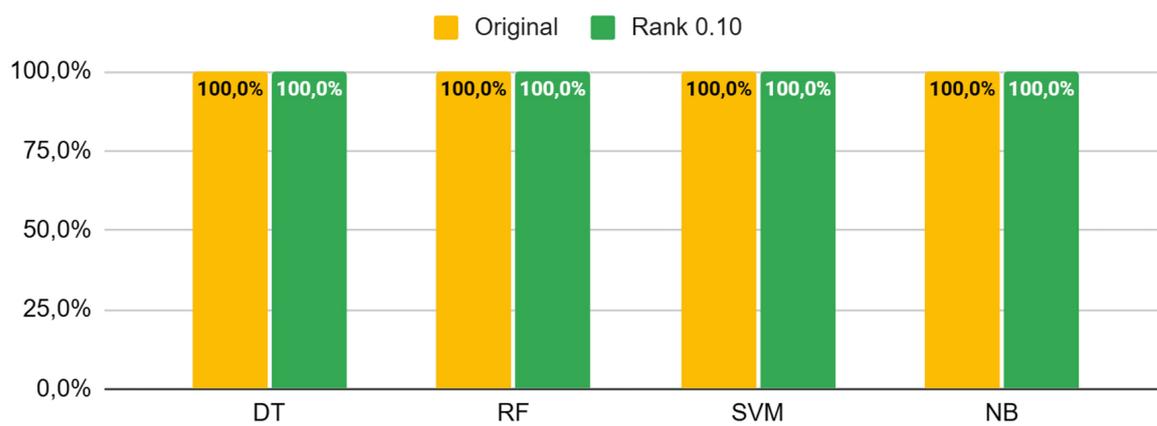


Figura 15 – Comparação entre os resultados obtidos nos algoritmos de ML avaliando o Banco Silva Desbalanceado.

Tabela 7 – Análise entre métodos de classificação de lesões na cavidade oral propostos na literatura.

Estudo	Extração de Características	Classificadores	Acurácia (%)
(ADEL et al., 2018)	ORB	SVM	92,6
(SILVA et al., 2022b)	<i>CNN features</i>	HOP	98,0
(DEIF et al., 2022)	<i>Learning feature</i>	XGBoost	96,3
(NEVES et al., 2023)	<i>Learning feature</i>	CNN	97,9
Método proposto (pior modelo)	CNN comprimida	DT	83,6
Método proposto (melhor modelo)	CNN comprimida	Softmax	100

oral no sentido de apresentar o uso de modelos de CNN comprimidos para auxiliar no processo de diagnóstico por especialistas.

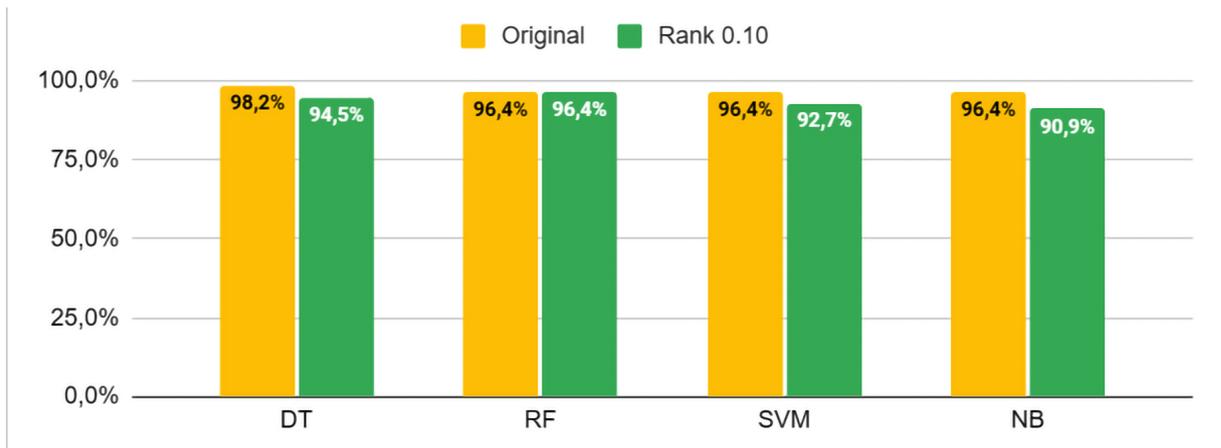


(a) Modelos ResNet-18 Original e Comprimido.

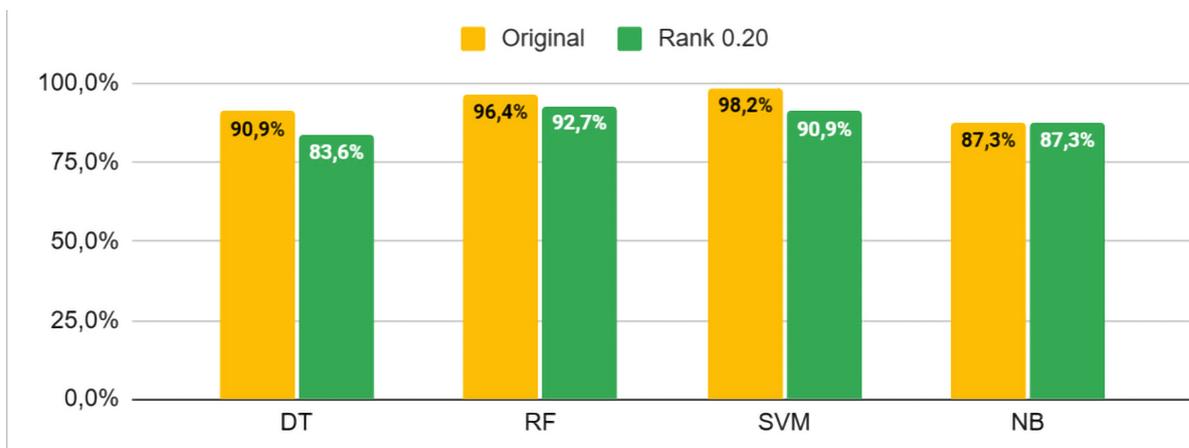


(b) Modelos ResNet-50 Original e Comprimido.

Figura 16 – Comparação entre os resultados obtidos nos algoritmos de ML avaliando o Banco Costa Balanceado.



(a) Modelos ResNet-18 Original e Comprimido.



(b) Modelos ResNet-50 Original e Comprimido.

Figura 17 – Comparação entre os resultados obtidos nos algoritmos de ML avaliando o Banco Costa Desbalanceado.

## 5 Conclusão

Este estudo investigou a aplicação da decomposição de Tucker nos tensores dos *kernels* das camadas convolucionais da arquitetura ResNet, com o objetivo de eficientemente comprimir o número de parâmetros do modelo original. Foram investigados essa estratégia sobre modelos da ResNet em que o número de camada é diferente. As operações de decomposição foram aplicadas somente às camadas convolucionais. Para realizar os experimentos, foram definidos 0,50, 0,40, 0,30, 0,20 e 0,10 em relação aos valores originais de cada dimensão em cada tensor. Esses valores foram usados nos modelos ResNet-18 e ResNet-50. Além da avaliação da classificação realizada pelas redes, foram examinados os descritores nas camadas da rede, conhecidos como mapas de características, por meio de algoritmos de aprendizado de máquina.

Os resultados em bancos de imagem balanceados, apresentados nas Tabelas 3 e 5, mostram que as redes conseguiram manter uma boa acurácia, mesmo quando decompostas com valores muito pequenos, representando uma compressão bastante significativa, com uma redução drástica no número total de parâmetros. Também foi avaliado o uso da decomposição em redes treinadas em bancos de imagem desbalanceados, onde foram obtido resultados com um nível de perda de acurácia maior, demonstrando que a técnica funciona melhor em bancos de imagem balanceados.

Ainda, neste trabalho avaliou-se a classificação utilizando algoritmos de ML treinados a partir das *features* extraídas das redes comprimidas. Novamente, os resultados mostraram-se interessantes para os conjuntos de imagem balanceados, mantendo uma alta acurácia na maioria dos casos. Em futuras pesquisas, serão investigados o impacto de normalização de corantes na acurácia, novas técnicas de decomposição e métodos mais sofisticados para a estimativa dos valores de *rank*.

## Referências

- ADEL, D.; MOUNIR, J.; EL-SHAFFEY, M.; ELDIN, Y. A.; MASRY, N. E.; ABDELRAOUF, A.; ELHAMID, I. S. A. Oral epithelial dysplasia computer aided diagnostic approach. In: **2018 13th International Conference on Computer Engineering and Systems (ICCES)**. Cairo, Egypt: IEEE, 2018. p. 313–318. Disponível em: <<https://doi.org/10.1109/ICCES.2018.8639452>>. Citado 2 vezes nas páginas 25 e 43.
- AL-HAIJA, Q. A.; ADEBANJO, A. Breast cancer diagnosis in histopathological images using resnet-50 convolutional neural network. In: **2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)**. [s.n.], 2020. p. 1–7. Disponível em: <<https://doi.org/10.1109/IEMTRONICS51293.2020.9216455>>. Citado na página 33.
- ATTY, A.; RIBEIRO, C.; ANTUNES, H.; SANTOS, R.; MIRANDA, ; LEITE, A. F.; MIGOWSKI, A.; GOLDEMBERG, D. C.; PIRES, F.; NETTO, J.; BUEXM, L.; ROMAñACH, M.; AGOSTINI, M.; LOURENÇO, S. **DIAGNÓSTICO PRECOCE DO CÂNCER DE BOCA**. Rio de Janeiro, RJ: Instituto Nacional de Câncer, 2022. ISBN 978-65-88517-20-8. Citado na página 13.
- BANERJEE, S.; PAL, M.; CHAKRABARTY, J.; PETIBOIS, C.; PAUL, R.; GIRI, A.; CHATTERJEE, J. Fourier-transform-infrared-spectroscopy based spectral-biomarker selection towards optimum diagnostic differentiation of oral leukoplakia and cancer. **Analytical and bioanalytical chemistry**, v. 407, 09 2015. Disponível em: <<https://doi.org/10.1007/s00216-015-8960-3>>. Citado na página 24.
- BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. In: **Proceedings of the Fifth Annual Workshop on Computational Learning Theory**. New York, NY, USA: Association for Computing Machinery, 1992. (COLT '92), p. 144–152. ISBN 089791497X. Disponível em: <<https://doi.org/10.1145/130385.130401>>. Citado na página 23.
- BREIMAN, L. Random forests. **Machine Learning**, Kluwer Academic Publishers, v. 45, n. 1, p. 5–32, 2001. ISSN 0885-6125. Disponível em: <<http://dx.doi.org/10.1023/A%3A1010933404324>>. Citado na página 23.
- CHEN, J.; CHEN, Z. Tensor decomposition based networks for nuclei segmentation and classification. **Electronics Letters**, v. 58, n. 25, p. 975–977, 2022. Disponível em: <<https://doi.org/10.1049/ell2.12668>>. Citado na página 10.
- CICHOCKI, A.; MANDIC, D.; LATHAUWER, L. D.; ZHOU, G.; ZHAO, Q.; CAIAFA, C.; PHAN, H. A. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. **IEEE Signal Processing Magazine**, v. 32, n. 2, p. 145–163, 2015. Disponível em: <<https://doi.org/10.1109/MSP.2013.2297439>>. Citado na página 22.
- CORTES, C.; VAPNIK, V. Support vector networks. **Machine Learning**, v. 20, p. 273–297, 1995. Disponível em: <<https://doi.org/10.1007/BF00994018>>. Citado na página 23.

DEIF, M. A.; ATTAR, H.; AMER, A.; ELHATY, I. A.; KHOSRAVI, M. R.; SOLYMAN, A. A. et al. Diagnosis of oral squamous cell carcinoma using deep neural networks and binary particle swarm optimization on histopathological images: an aiomt approach. **Computational Intelligence and Neuroscience**, Hindawi, v. 2022, 2022. Disponível em: <<https://doi.org/10.1155/2022/6364102>>. Citado na página 43.

GABOR, M.; ZDUNEK, R. Convolutional neural network compression via tensor-train decomposition on permuted weight tensor with automatic rank determination. In: GROEN, D.; MULATIER, C. de; PASZYNSKI, M.; KRZHIZHANOVSKAYA, V. V.; DONGARRA, J. J.; SLOOT, P. M. A. (Ed.). **Computational Science – ICCS 2022**. Cham: Springer International Publishing, 2022. p. 654–667. ISBN 978-3-031-08757-8. Disponível em: <[https://doi.org/10.1007/978-3-031-08757-8\\_54](https://doi.org/10.1007/978-3-031-08757-8_54)>. Citado na página 27.

\_\_\_\_\_. Compressing convolutional neural networks with hierarchical tucker-2 decomposition. **Applied Soft Computing**, v. 132, p. 109856, 2023. ISSN 1568-4946. Disponível em: <<https://doi.org/10.1016/j.asoc.2022.109856>>. Citado na página 28.

GONZALEZ, R.; WOODS, R. **Processamento Digital De Imagens**. ADDISON WESLEY BRA, 2009. ISBN 9788576054016. Disponível em: <<https://books.google.com.br/books?id=r5f0RgAACAAJ>>. Citado na página 15.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, United States of America: MIT Press, 2016. <<http://www.deeplearningbook.org>>. Citado 3 vezes nas páginas 19, 20 e 21.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. Hamilton, Canada: Prentice Hall, 1999. ISBN 81-7808-300-0. Citado 2 vezes nas páginas 18 e 19.

HAYKIN, S. S. **Neural networks and learning machines**. Third. Upper Saddle River, NJ: Pearson Education, 2009. ISBN 0-13-147139-2. Citado 4 vezes nas páginas 17, 18, 19 e 20.

HE, K.; GKIOXARI, G.; DOLLÁR, P.; GIRSHICK, R. **Mask R-CNN**. 2018. Disponível em: <<https://doi.org/10.1109/ICCV.2017.322>>. Citado na página 25.

HE, K.; ZHANG, X.; REN, S.; SUN, J. **Deep Residual Learning for Image Recognition**. 2015. Disponível em: <<https://doi.org/10.1109/CVPR.2016.90>>. Citado na página 33.

INCA. **Falando sobre o câncer da boca**. Rio de Janeiro, RJ: Engenho e Arte, 2002. ISBN 85-7318-060-9. Disponível em: <[https://bvsm.sau.gov.br/bvs/publicacoes/falando\\_sobre\\_cancer\\_boca.pdf](https://bvsm.sau.gov.br/bvs/publicacoes/falando_sobre_cancer_boca.pdf)>. Citado 2 vezes nas páginas 10 e 13.

KAUSHAL, C.; BHAT, S.; KOUNDAL, D.; SINGLA, A. Recent trends in computer assisted diagnosis (cad) system for breast cancer diagnosis using histopathological images. **IRBM**, v. 40, n. 4, p. 211–227, 2019. ISSN 1959-0318. Disponível em: <<https://doi.org/10.1016/j.irbm.2019.06.001>>. Citado 2 vezes nas páginas 15 e 16.

KIM, Y.; PARK, E.; YOO, S.; CHOI, T.; YANG, L.; SHIN, D. Compression of deep convolutional neural networks for fast and low power mobile applications. In: BENGIO, Y.; LECUN, Y. (Ed.). **4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track**

**Proceedings**. ICLR, 2016. Disponível em: <<http://arxiv.org/abs/1511.06530>>. Citado 5 vezes nas páginas 10, 11, 26, 29 e 34.

KOLDA, T. G.; BADER, B. W. Tensor decompositions and applications. **SIAM Review**, SIAM, v. 51, n. 3, p. 455–500, 2009. ISSN 00361445. Disponível em: <<http://dx.doi.org/10.1137/07070111X>>. Citado na página 34.

KUMAR, V. **Robbins & cotran - patologia bases patológicas das doenças 8a edição**. Elsevier Brasil, 2010. ISBN 9788535246339. Disponível em: <[https://books.google.com.br/books?id=2O\\\_jYLsNxrYC](https://books.google.com.br/books?id=2O\_jYLsNxrYC)>. Citado 2 vezes nas páginas 10 e 13.

KUMAR, Y.; KOUL, A.; SINGLA, R.; IJAZ, M. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. **Journal of Ambient Intelligence and Humanized Computing**, Springer Verlag, 2022. ISSN 1868-5137. Funding Information: This research work was supported by Sejong University research fund. Yogesh Kumar and Muhammad Fazal Ijaz contributed equally to this work and are first co-authors. Publisher Copyright: © 2021, The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature. Disponível em: <<https://doi.org/10.1007/s12652-021-03612-z>>. Citado na página 15.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998. Disponível em: <<https://doi.org/10.1109/5.726791>>. Citado na página 19.

LIMA, F.; OLIVEIRA, J.; COSTA, L.; MARTINS, L.; SANTOS, M.; CANCELA, M. de C.; MIGOWSKI, A.; SCHILITZ, A.; CORRÊA, F.; OLIVEIRA, I.; SANTOS, J.; SOUZA, L.; ALMEIDA, L.; GUIMARÃES, M.; OTERO, U. **Estimativa 2023 : incidência de câncer no Brasil**. Rio de Janeiro, RJ: Instituto Nacional de Câncer, 2023. ISBN 978-65-88517-10-9. Citado na página 11.

LISBOA, P. J. G. A review of evidence of health benefit from artificial neural networks in medical intervention. **Neural networks : the official journal of the International Neural Network Society**, v. 15, n. 1, p. 11–39, January 2002. ISSN 0893-6080. Disponível em: <[https://doi.org/10.1016/s0893-6080\(01\)00111-3](https://doi.org/10.1016/s0893-6080(01)00111-3)>. Citado 2 vezes nas páginas 10 e 15.

LIU, Y.; NG, M. K. Deep neural network compression by tucker decomposition with nonlinear response. **Knowledge-Based Systems**, v. 241, p. 108171, 2022. ISSN 0950-7051. Disponível em: <<https://doi.org/10.1016/j.knosys.2022.108171>>. Citado 2 vezes nas páginas 28 e 29.

LUMERMAN, H.; FREEDMAN, P.; KERPEL, S. Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. **Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology**, v. 79, n. 3, p. 321–329, 1995. ISSN 1079-2104. Disponível em: <[https://doi.org/10.1016/S1079-2104\(05\)80226-4](https://doi.org/10.1016/S1079-2104(05)80226-4)>. Citado 3 vezes nas páginas 13, 14 e 31.

MITCHELL, T. M. **Machine learning**. [S.l.]: McGraw-hill New York, 1997. v. 1. Citado na página 23.

- NEVES, L. A.; MARTINEZ, J. M. C.; LONGO, L. H. d. C.; ROBERTO, G. F.; TOSTA, T. A. A.; FARIA, P. R. d.; LOYOLA, A. M.; CARDOSO, S. V.; SILVA, A. B.; NASCIMENTO, M. Z. d.; ROZENDO, G. B. Classification of the images via cnn models with xai approaches, deepdream representations and multiple classifiers. In: **International Conference on Enterprise Information Systems - ICEIS**. SciTePress, 2023. Disponível em: <<https://doi.org/10.5220/0011839400003467>>. Citado na página 43.
- OLIVEIRA, R. F. d. **Inteligência Artificial**. Londrina, PR: Editora e Distribuidora Educacional S.A, 2018. ISBN 978-85-522-1141-9. Citado 2 vezes nas páginas 17 e 18.
- PAIXÃO, G. M. d. M.; SANTOS, B. C.; ARAUJO, R. M. d.; RIBEIRO, M. H.; MORAES, J. L. d.; RIBEIRO, A. L. Machine learning in medicine: Review and applicability. **Arquivos brasileiros de cardiologia**, v. 118, n. 1, p. 95—102, January 2022. ISSN 0066-782X. Disponível em: <<https://doi.org/10.36660/abc.20200596>>. Citado 2 vezes nas páginas 10 e 14.
- PHAN, A.-H.; SOBOLEV, K.; SOZYKIN, K.; ERMILOV, D.; GUSAK, J.; TICHAVSKY, P.; GLUKHOV, V.; OSELEDETS, I.; CICHOCKI, A. **Stable Low-rank Tensor Decomposition for Compression of Convolutional Neural Network**. 2020. Disponível em: <[https://doi.org/10.1007/978-3-030-58526-6\\_31](https://doi.org/10.1007/978-3-030-58526-6_31)>. Citado na página 27.
- QUINLAN, J. R. Induction of decision trees. **Machine Learning**, v. 1, p. 81–106, 1986. Disponível em: <<https://doi.org/10.1007/BF00116251>>. Citado na página 22.
- ROCHA, K. da; BERMUDEZ, J. C. M.; RIVERO, E. R. C.; COSTA, M. H. **A Pathology-Based Machine Learning Method to Assist in Epithelial Dysplasia Diagnosis**. 2022. Citado na página 26.
- SADRI, G.; MAHJUB, H. Tobacco smoking and oral cancer: A meta-analysis. **Journal of research in health sciences**, v. 7, p. 18–23, 07 2007. Citado na página 13.
- SARWINDA, D.; PARADISA, R. H.; BUSTAMAM, A.; ANGGIA, P. Deep learning in image classification using residual network (resnet) variants for detection of colorectal cancer. **Procedia Computer Science**, v. 179, p. 423–431, 2021. ISSN 1877-0509. 5th International Conference on Computer Science and Computational Intelligence 2020. Disponível em: <<https://doi.org/10.1016/j.procs.2021.01.025>>. Citado na página 33.
- SIDIROPOULOS, N. D.; LATHAUWER, L. D.; FU, X.; HUANG, K.; PAPALEXAKIS, E. E.; FALOUTSOS, C. Tensor decomposition for signal processing and machine learning. **IEEE Transactions on Signal Processing**, v. 65, n. 13, p. 3551–3582, 2017. Disponível em: <<https://doi.org/10.1109/TSP.2017.2690524>>. Citado na página 21.
- SILVA, A. B. Métodos computacionais para análise e classificação de displasias em imagens da cavidade bucal. 2019. Disponível em: <<http://dx.doi.org/10.14393/ufu.di.2019.2390>>. Citado na página 31.
- 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.; do Nascimento, M. Z. Computational analysis of histological images from hematoxylin and eosin-stained oral epithelial dysplasia tissue sections. **Expert Systems with Applications**, v. 193, p. 116456, 2022.

ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417421017401>>. Citado 3 vezes nas páginas 14, 25 e 34.

SILVA, A. B.; OLIVEIRA, C. I. D.; PEREIRA, D. C.; TOSTA, T. A. A.; MARTINS, A. S.; LOYOLA, A. M.; CARDOSO, S. V.; FARIA, P. R. D.; NEVES, L. A.; NASCIMENTO, M. Z. D. Assessment of the association of deep features with a polynomial algorithm for automated oral epithelial dysplasia grading. In: **2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. [s.n.], 2022. v. 1, p. 264–269. Disponível em: <<https://doi.org/10.1109/SIBGRAPI55357.2022.9991758>>. Citado na página 43.

TALO, M. Convolutional neural networks for multi-class histopathology image classification. **ArXiv**, abs/1903.10035, 2019. Disponível em: <<https://api.semanticscholar.org/CorpusID:85497281>>. Citado na página 33.

WARNAKULASURIYA, S.; REIBEL, J.; BOUQUOT, J.; DABELSTEEN, E. Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. **Journal of Oral Pathology & Medicine**, v. 37, n. 3, p. 127–133, 2008. Disponível em: <<https://doi.org/10.1111/j.1600-0714.2007.00584.x>>. Citado 3 vezes nas páginas 10, 13 e 14.