



Universidade Federal do ABC  
Centro de Matemática, Computação e Cognição  
Bacharelado em Ciência da Computação

# **Detecção e Classificação Automática de Osteoartrite de Joelho em Radiografias Utilizando Redes Neurais Convolucionais**

**Guilhere de Sousa Santos**

**Santo André - SP, 22 de agosto de 2024**



Guilhere de Sousa Santos

# **Deteccção e Classificação Automática de Osteoartrite de Joelho em Radiografias Utilizando Redes Neurais Convolucionais**

**Projeto de Graduação** apresentado ao Programa de Graduação em Ciência da Computação (área de concentração: Visão Computacional), como parte dos requisitos necessários para a obtenção do título de Bacharel em Ciência da Computação.

Universidade Federal do ABC – UFABC

Centro de Matemática, Computação e Cognição

Bacharelado em Ciência da Computação

Orientador: Hugo Puertas de Araújo

Santo André - SP

22 de agosto de 2024



# Lista de abreviaturas e siglas

OA	Osteoarthritis
KOA	Knee Osteoarthritis
KL	Kellgren/Lawrence
CT	Tomografia Computadorizada
MRI	Magnetic Resonance Imaging
CNN	Convolutional Neural Network
ViT	Visual Transformer



# Sumário

	<b>Introdução</b>	<b>1</b>
<b>1</b>	<b>METODOLOGIA</b>	<b>5</b>
<b>1.1</b>	<b>Coleta de dados</b>	<b>5</b>
<b>1.2</b>	<b>Pré-processamento das imagens</b>	<b>5</b>
1.2.1	Normalização	6
1.2.2	Equalização	6
1.2.3	Filtragem e suavização	6
1.2.4	Aumento de dados	7
<b>1.3</b>	<b>Arquitetura do modelo de Rede Neural Convolucional</b>	<b>7</b>
1.3.1	ResNet (Residual Network)	7
1.3.2	VGG (Visual Geometry Group Network)	8
1.3.3	DenseNet (Densely Connected Convolutional Networks)	8
1.3.4	Inception (GoogLeNet)	8
<b>1.4</b>	<b>Arquitetura do modelo de Vision Transformer</b>	<b>8</b>
1.4.1	ViT-B/16	9
1.4.2	DeiT (Data-efficient Image Transformer)	9
1.4.3	Swin Transformer (Shifted Window Transformer)	9
	<b>REFERÊNCIAS</b>	<b>11</b>





# Introdução

A osteoartrite (OA) é uma forma muito comum de doença articular, definida como uma condição degenerativa que se inicia nas articulações e afeta principalmente a cartilagem, o revestimento articular e os ligamentos (SARDIM; PRADO; PINFILDI, 2020), resultando em sintomas de dor, rigidez e mobilidade articular limitada (PACCA et al., 2018). Tais fatores podem comprometer significativamente a qualidade de vida, especialmente em idosos e indivíduos obesos (PACCA et al., 2022). A OA é altamente prevalente, sendo uma das principais causas de incapacidade no mundo, com grande incidência em articulações como joelhos e quadris, afetando uma em cada sete pessoas globalmente (KELLGREN; LAWRENCE, 1957). De acordo com um estudo do World Health Organization (WHO), em 2023, estimava-se a prevalência global da OA de joelho em 365 milhões de indivíduos, com maior predominância em pessoas idosas e mulheres, com cerca de 70% e 60%, respectivamente [TODO: colocar a fonte].

Terapias farmacêuticas têm sido aplicadas a pacientes diagnosticados com OA de joelho com o objetivo de reduzir os sintomas de dor, uma vez que não existem medicamentos capazes de retardar o seu desenvolvimento. No entanto, a progressão da doença pode ser prevenida com um diagnóstico precoce, ou seja, nos estágios iniciais em que a OA de joelho ainda é reversível (KANAMOTO et al., 2020). A medicina comumente avalia a severidade da doença através dos graus de Kellgren/Lawrence (KL), que categoriza a doença em cinco níveis de progressão: 0 (saudável), 1 (duvidoso), 2 (mínimo), 3 (moderado) e 4 (severo), dependendo da experiência e cuidado médico na interpretação das radiografias (KELLGREN; LAWRENCE, 1957). Isso pode levar a inconsistências entre o grau previsto e o grau real da OA de joelho, devido às mínimas diferenças entre os estágios adjacentes da doença (MOHAMMED et al., 2023). Estudos indicam que procedimentos como artroscopia são invasivos e podem causar complicações (SARAEV et al., 2020), enquanto técnicas como tomografia computadorizada (CT) e ressonância magnética (MRI) também são usadas, mas o diagnóstico pode ser impreciso por falta de experiência do profissional (ALSHAMRANI et al., 2023). Esses desafios têm impulsionado estudos sobre sistemas automáticos de detecção e classificação da OA de joelho.

Nos últimos anos, muitas áreas têm visto a introdução de sistemas de inteligência artificial (IA) para executar tarefas que eram realizadas de forma manual, incluindo na área da medicina para o diagnóstico de patologias, por exemplo. Avanços recentes em técnicas de aprendizado de máquina no campo da saúde levaram a uma aceleração no diagnóstico de diversas doenças, incluindo a OA de joelho (MOHAMMED et al., 2023). O uso de modelos de aprendizado profundo baseados em redes neurais convolucionais (RNCs) tem ganhado espaço no que tange tarefas relacionadas a visão computacional (TARIQ;

SUHAIL; NAWAZ, 2023). Porém, isso só foi possível após a introdução de novas técnicas para treinar redes profundas em paralelo com avanços a nível de hardware (LITJENS et al., 2017). Aprendizado por transferência também é amplamente utilizado para reduzir uso de recursos computacionais para tarefas que já são executadas por modelos existentes, como as redes residuais (ResNet), Visual Geometry Group (VGG) e as redes densamente conectadas (DenseNet) (TARIQ; SUHAIL; NAWAZ, 2023). Enquanto o uso de RNCs tem se mostrado útil em soluções de detecção em imagens médicas, a operação de convolução limita o relacionamento entre pixels distantes numa imagem. Para tanto, a habilidade de codificar dependências de longo alcance tem sido possível graças às arquiteturas de aprendizado profundas baseadas em atenção, como o Vision Transformer (ViT). Tais modelos de ViT têm sido empregados para várias tarefas, incluindo classificação e detecção de objetos (SHAMSHAD et al., 2023).

A relevância desta pesquisa reside na necessidade de aprimorar o processo de diagnóstico da osteoartrite de joelho, uma doença que afeta milhões de pessoas em todo o mundo e cuja detecção precoce é crucial para retardar sua progressão. O diagnóstico manual, feito por radiologistas, muitas vezes é subjetivo e suscetível a erros, o que pode levar a diagnósticos tardios ou incorretos. A aplicação de RNCs oferece uma solução promissora para automatizar esse processo, proporcionando uma avaliação mais precisa e eficiente a partir de radiografias. Essa automatização pode reduzir a carga dos profissionais de saúde e aumentar a acessibilidade de diagnósticos mais rápidos e confiáveis. Além disso, a comparação entre arquiteturas de RNCs e modelos baseados em transformers é relevante para identificar qual abordagem oferece melhor desempenho na classificação da severidade da osteoartrite. Embora as RNCs sejam amplamente utilizadas em tarefas de visão computacional, arquiteturas de transformers têm se mostrado eficazes em diversas áreas, incluindo na medicina para o diagnóstico de doenças.

O objetivo deste trabalho consiste em realizar uma comparação entre as métricas de performance e eficiência computacional de RNCs e modelos de ViTs na tarefa de detecção e classificação da OA de joelho seguindo a escala de Kellgren/Lawrence a partir de radiografias, com o intuito de identificar qual abordagem é mais adequada para uso em diagnósticos clínicos. Para atingir esse objetivo, será necessário estudar as modelos de RNCs e ViTs e propor uma arquitetura capaz de solucionar o problema. Em seguida, deverá ser feito o treinamento dos modelos e, por fim, será realizada uma análise detalhada das métricas de performance, incluindo acurácia, tempo de processamento e consumo de recursos, permitindo uma avaliação comparativa das duas abordagens.

A metodologia adotada envolve o uso de técnicas de pré-processamento de imagens para reduzir ruído das radiografias, melhorar contraste e ampliar o dataset para melhorar a performance e evitar o problema de *overfitting*. Em seguida, diferentes arquiteturas de RNCs serão treinadas usando a estratégia de *transfer learning*, e seus resultados comparados

com os de um modelo de vision transformer treinado para a mesma tarefa.



# 1 Metodologia

Esta seção descreve a metodologia proposta para atingir os objetivos da pesquisa.

## 1.1 Coleta de dados

A escolha e coleta dos dados é a primeira tarefa a ser realizada quando o objetivo é treinar um modelo de aprendizado profundo, incluindo redes neurais artificiais e vision transformers. Um conjunto de dados adequado é essencial para que o modelo tenha uma boa performance e seja útil para se tornar uma ferramenta de suporte no diagnóstico de OA de joelho. O conjunto de dados foi obtido a partir da plataforma Kaggle [TODO: fonte], uma fonte amplamente reconhecida por fornecer dados de alta qualidade e de domínio público para estudos acadêmicos e projetos de aprendizado de máquina. O conjunto de dados escolhido é baseado na Osteoarthritis Initiative (OAI) [TODO: fonte], um estudo observacional multicêntrico de dez anos de homens e mulheres, patrocinado pelo National Institutes of Health (NIH), com o objetivo de permitir uma melhor compreensão da prevenção e tratamento da osteoartrite de joelho. Este conjunto contém radiografias de joelhos, juntamente com suas respectivas classificações de severidade da OA, conforme o sistema de Kellgren/Lawrence. Este dataset foi selecionado por sua relevância na plataforma, fornecendo uma base sólida para o treinamento dos modelos de RNCs e ViTs propostos nesta pesquisa. A [Tabela 1](#) detalha algumas propriedades do dataset.

O conjunto de dados contém quatro pastas nomeadas “auto\_test”, “test”, “train” e “val”, cada uma contendo as subpastas com imagens 224x224 representando cada um dos graus de KL. O dataset foi dividido entre dados de treino, teste e validação, com uma proporção de 7:2:1. O conjunto de treino é usado para treinar os modelos e consiste na maior proporção de imagens. O conjunto de validação é usado para ajustar os hiperparâmetros do modelo e monitorar o seu desempenho, enquanto o conjunto de teste é usado após o treinamento completo do modelo, para medir o desempenho final e verificar sua capacidade de generalização em dados completamente novos.

## 1.2 Pré-processamento das imagens

O pré-processamento de imagens de raio-X é crucial para melhorar a qualidade e facilitar a análise automatizada pelos modelos. Para isso, algumas técnicas devem ser utilizadas, incluindo:

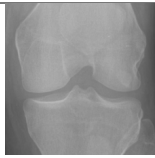




Classe KL	Número de Imagens	Exemplo de Imagem
0 (saudável)	3857	
1 (duvidoso)	1770	
2 (mínimo)	2578	
3 (moderado)	1286	
4 (severo)	295	

Tabela 1 – Tabela do conjunto de dados escolhido.

### 1.2.1 Normalização

A normalização de dados visa ajustar os valores para um intervalo padrão, melhorando a consistência dos dados e a eficiência dos modelos treinados. Para as radiografias, os pixels devem ter seus valores transformados para o intervalo entre 0 e 1.

### 1.2.2 Equalização

A equalização busca melhorar o contraste e a visibilidade dos detalhes em uma imagem. O objetivo da equalização é redistribuir os níveis de cinza para que todos os valores de intensidade apareçam com uma frequência mais uniforme. Isso é realizado calculando o histograma acumulado da radiografia original e utilizando-o para redistribuir os valores de pixel.

### 1.2.3 Filtragem e suavização

As técnicas de filtragem e suavização visam reduzir o ruído e melhorar a qualidade visual das imagens. Para isso, será utilizado o filtro gaussiano, onde cada pixel tem o seu valor substituído pelo valor ponderado da média dos pixels vizinhos, onde os pesos são determinados pela função gaussiana.

### 1.2.4 Aumento de dados

A ideia desta técnica é expandir artificialmente o tamanho e a variabilidade de um conjunto de dados, principalmente quando o volume de dados disponível é limitado. Isso torna os modelos mais robustos e genéricos, prevenindo *overfitting* e melhorar o desempenho em dados novos. As técnicas de aumento de dados que serão utilizadas nas radiografias são: rotação, escalonamento e reflexão (espelhamento) horizontal.

## 1.3 Arquitetura do modelo de Rede Neural Convolutacional

As redes neurais convolucionais possuem um papel muito relevante no contexto de inteligência artificial, especialmente em tarefas de visão computacional devido à sua capacidade de extrair características relevantes de imagens de forma automática, sem qualquer intervenção manual. Sua arquitetura é especialmente eficaz para reconhecer e classificar objetos em imagens complexas, inclusive em radiografias, com o intuito de auxiliar no processo de diagnóstico médico. As RNCs conseguem identificar variações sutis que podem estar associadas a condições patológicas, como é o caso da osteoartrite de joelho, onde as variações entre os graus de KL reside no espaçamento articular da junção do joelho.

Fazer o treinamento de uma RNC sem nenhum conhecimento prévio do modelo é custoso em termos de quantidade de dados necessário, consumo de recursos computacionais e tempo. Para resolver este problema, o uso de *transfer learning* é essencial, pois permite aproveitar modelos já treinados em grandes conjuntos de dados genéricos, como o ImageNet, e adaptá-los para o conjunto de dados específico para o problema. Ao utilizar o *transfer learning*, as primeiras camadas do modelo, que capturam características gerais da imagem, são congeladas, enquanto as camadas finais são ajustadas para a tarefa específica, tal processo é chamado de *fine-tuning*. Isso economiza tempo e recursos computacionais e aumenta a eficácia do treinamento, resultando em modelos que podem fornecer diagnósticos precisos mesmo com volumes menores de dados disponíveis. Nos últimos anos, algumas arquiteturas performaram muito bem em algumas tarefas, como por exemplo a ResNet, VGG, GoogLeNet e DenseNet. Para esta pesquisa, serão trabalhadas os seguintes modelos de RNC:

### 1.3.1 ResNet (Residual Network)

A ResNet é uma arquitetura amplamente utilizada em tarefas de classificação de imagens devido à sua capacidade de treinar redes profundas sem problemas de desaparecimento de gradiente. A inovação da ResNet está em seus blocos residuais, que introduzem conexões de atalho para permitir que os gradientes fluam melhor durante o treinamento. Isso torna a ResNet altamente eficiente para tarefas de classificação de imagens médicas.

Para este trabalho, serão treinados os modelos ResNet34, ResNet50 e ResNet101, que oferecem um bom equilíbrio entre profundidade e performance.

### 1.3.2 VGG (Visual Geometry Group Network)

O VGG é um modelo mais simples comparado ao ResNet, mas ainda é muito eficaz. Ele se destaca por usar camadas convolucionais de pequenos filtros (3x3) empilhadas seguidas por camadas de pooling. Embora o VGG tenha mais parâmetros que modelos mais modernos, sua estrutura é eficaz para capturar detalhes visuais em imagens médicas. O VGG16 e VGG19 serão utilizados nesta pesquisa.

### 1.3.3 DenseNet (Densely Connected Convolutional Networks)

O DenseNet utiliza conexões densamente conectadas, onde cada camada recebe entradas de todas as camadas anteriores. Isso promove um fluxo eficiente de gradientes e incentiva o reuso de características aprendidas, o que pode ser muito útil nas radiografias de osteoartrite de joelho, onde detalhes finos precisam ser capturados, especialmente na diferenciação entre graus de KL adjacentes. Os modelos do DenseNet121 e DenseNet169 serão as opções para este trabalho.

### 1.3.4 Inception (GoogLeNet)

A rede Inception, também chamada de GoogLeNet, é conhecida por seu uso de módulos Inception, que permitem que a rede aprenda de forma mais eficiente ao explorar convoluções de diferentes tamanhos em paralelo. A habilidade da Inception de capturar informações em várias escalas pode ser especialmente útil ao lidar com imagens médicas de diferentes resoluções. O Inception-v3 é uma versão mais moderna e, portanto, será utilizada nesta pesquisa.

## 1.4 Arquitetura do modelo de Vision Transformer

O modelo de Vision Transformer tem se destacado como uma abordagem poderosa para tarefas de visão computacional devido à sua capacidade de capturar relações globais em imagens através do mecanismo de atenção. Essa abordagem permite que os modelos de ViTs superem as limitações das RNCs, estas que focam mais em características locais da imagem. Tal capacidade é particularmente relevante para o diagnóstico de patologias em imagens médicas, incluindo radiografias, onde o modelo é capaz de processar toda a imagem simultaneamente, associando partes distantes e próximas com igual relevância. Além disso, os ViTs também se beneficiam do *transfer learning*, permitindo que os modelos sejam treinados de forma eficiente em conjuntos de dados limitados. Para esta pesquisa



serão treinados alguns modelos de ViTs para a tarefa de classificação da OA de joelho, como segue:

### 1.4.1 ViT-B/16

O ViT-B/16 é uma das primeiras variantes da arquitetura Vision Transformer, onde "B" representa o modelo base (base model) e "16" refere-se ao tamanho do patch em que a imagem é dividida (16x16 pixels). O ViT-B/16 recebe uma imagem e a divide em patches, tratando cada patch como um "token" semelhante ao processamento de palavras em texto nos transformers tradicionais. O modelo usa um mecanismo de atenção para processar os tokens de maneira global, capturando interdependências entre diferentes regiões da radiografia. Essa abordagem permite que o ViT-B/16 compreenda melhor a estrutura geral da imagem, identificando padrões que podem se estender por grandes áreas da mesma. Este modelo pode ser eficaz para a tarefa de classificação da OA de joelho ao capturar relações no espaçamento articular da divisão do joelho.

### 1.4.2 DeiT (Data-efficient Image Transformer)

O DeiT é uma versão otimizada dos ViTs, projetada para melhorar a eficiência no uso de dados. Enquanto os ViTs originais, como o ViT-B/16, geralmente precisam de grandes quantidades de dados para atingir um bom desempenho, o DeiT foi projetado para ser treinado em conjunto de dados reduzidos. Isso acontece devido à técnicas do *distillation token*, que permite ao modelo aprender a partir de um "professor" (modelo mais simples), aumentando a eficiência do treinamento. Este modelo pode ser particularmente útil na tarefa de classificação da OA de joelho, podendo ser um importante fator ao comparar com outros modelos de ViTs e RNCs.

### 1.4.3 Swin Transformer (Shifted Window Transformer)

O Swin Transformer é uma arquitetura de ViT que introduz uma abordagem inovadora ao utilizar janelas deslizantes, ou *sliding windows*, para aplicar a atenção. Em vez de processar toda a imagem como uma sequência de patches globalmente, o Swin Transformer aplica a atenção dentro de pequenas janelas locais, de forma hierárquica, permitindo que o modelo mantenha a eficiência computacional e ainda capture detalhes locais e globais. Conforme o modelo avança pelas camadas, as janelas se expandem e se deslocam, permitindo que o Swin Transformer agregue contexto global ao longo do processamento. Essa estrutura hierárquica é particularmente eficaz para imagens de alta resolução, como as radiografias, onde há muitos detalhes importantes em diferentes escalas. Além disso, o Swin Transformer pode ser facilmente escalado para diferentes tamanhos de imagens e é altamente eficiente em termos de uso de memória e poder computacional, sendo uma escolha apropriada para a tarefa de classificação da OA de joelho.



# Referências

- ALSHAMRANI, H. A. et al. Osteo-net: An automated system for predicting knee osteoarthritis from x-ray images using transfer-learning-based neural networks approach. *Healthcare (Switzerland)*, v. 11, 2023. ISSN 22279032. Citado na página 1.
- KANAMOTO, T. et al. *Significance and definition of early knee osteoarthritis*. 2020. Citado na página 1.
- KELLGREN, J. H.; LAWRENCE, J. S. Radiological assessment of osteo-arthritis. *Annals of the rheumatic diseases*, v. 16, 1957. ISSN 00034967. Citado na página 1.
- LITJENS, G. et al. *A survey on deep learning in medical image analysis*. 2017. Citado na página 2.
- MOHAMMED, A. S. et al. Knee osteoarthritis detection and severity classification using residual neural networks on preprocessed x-ray images. *Diagnostics*, v. 13, 2023. ISSN 20754418. Citado na página 1.
- PACCA, D. M. et al. Prevalência de dor articular e osteoartrite na população obesa brasileira. *ABCD. Arquivos Brasileiros de Cirurgia Digestiva (São Paulo)*, v. 31, 2018. ISSN 2317-6326. Citado na página 1.
- PACCA, D. M. et al. Desenvolvimento e aplicação de rede neural convolucional para o diagnóstico de osteoartrite de joelho. *Revista CPAQV - Centro de Pesquisas Avançadas em Qualidade de Vida*, v. 15, 2022. Disponível em: <<https://revista.cpaqv.org/index.php/CPAQV/article/view/1079>>. Citado na página 1.
- SARAEV, A. V. et al. Arthroscopy for knee osteoarthritis in the xxi century: a systematic review of current high quality researches and guidelines of professional societies. *Traumatology and Orthopedics of Russia*, v. 26, 2020. ISSN 2311-2905. Citado na página 1.
- SARDIM, A. C.; PRADO, R. P.; PINFILDI, C. E. Efeito da fotobiomodulação associada a exercícios na dor e na funcionalidade de pacientes com osteoartrite de joelho: estudo-piloto. *Fisioterapia e Pesquisa*, v. 27, 2020. ISSN 1809-2950. Citado na página 1.
- SHAMSHAD, F. et al. *Transformers in medical imaging: A survey*. 2023. Citado na página 2.
- TARIQ, T.; SUHAIL, Z.; NAWAZ, Z. Knee osteoarthritis detection and classification using x-rays. *IEEE Access*, v. 11, 2023. ISSN 21693536. Citado na página 2.