

Classificação de Graus de Osteoartrite de Joelho em Raio-X via *EfficientNetV2* com Atenção por Canal e Otimização Bayesiana

Walberto M. Santos¹, Marcos R. A. Amorim¹,
João O. B. Diniz^{1,2}, Neilson P. Ribeiro^{1,2},
Geraldo B. Júnior¹, João Dallyson S. Almeida¹

¹Núcleo de Computação Aplicada — Universidade Federal do Maranhão (UFMA)
Caixa Postal 65.085–580 — São Luís — MA — Brasil

²Fábrica de Inovação - Instituto Federal do Maranhão (IFMA)
Caixa Postal 65.940–000 — Grajaú — MA — Brasil

walberto.marques@ufma.br

Abstract. *Knee osteoarthritis (KOA) is one of the main causes of pain and functional disability, and is generally classified using the Kellgren-Lawrence (KL) scale. Differentiating between adjacent grades of this scale represents a challenge due to the similarity of joint alterations. The method proposed in this work aims to classify the severity of KOA using the EfficientNetV2B0 architecture combined with a channel attention mechanism and Bayesian optimization for hyperparameter tuning. The obtained results demonstrate promising performance, achieving an accuracy of 77.58%, F1-score of 76.98%, and Kappa coefficient of 70.48%. The results of the method indicate potential to assist specialists in the diagnosis and evaluation of KOA.*

Resumo. *A osteoartrite de joelho (OAJ) é uma das principais causas de dor e incapacidade funcional, sendo geralmente classificada via escala de Kellgren-Lawrence (KL). A diferenciação entre graus adjacentes dessa escala representa um desafio pela similaridade das alterações articulares. O método proposto neste trabalho visa classificar da severidade da OAJ utilizando a arquitetura EfficientNetV2B0 combinada com um mecanismo de atenção por canal e otimização Bayesiana para ajuste de hiperparâmetros. Os resultados obtidos demonstram desempenho promissor, alcançando acurácia de 77,58%, F1-score de 76,98% e coeficiente Kappa de 70,48%. Os resultados do método indicam potencial para auxiliar especialistas no diagnóstico e na avaliação da OAJ.*

1. Introdução

A osteoartrite de joelho (OAJ) é uma das principais causas de dor crônica e limitação funcional em adultos e idosos. Estima-se que a osteoartrite afete quase 528 milhões de pessoas globalmente, com o joelho correspondendo entre 60% a 85% desses casos, uma incidência que segue aumentando impulsionada pelo envelhecimento populacional e pela crescente prevalência de obesidade [Li et al. 2024, Langworthy et al. 2024]. Esse cenário configura um desafio crescente para os sistemas de saúde, demandando ferramentas diagnósticas mais eficazes.

O diagnóstico da OAJ é realizado predominantemente por meio de radiografias convencionais, utilizando a escala de Kellgren-Lawrence (KL) [Kellgren and Lawrence 1957], que estratifica a doença em cinco graus com base em achados como estreitamento articular e formação de osteófitos. Entretanto, a interpretação radiográfica é subjetiva e apresenta expressiva variabilidade interobservador, especialmente nos graus intermediários KL 1 e KL 2, correspondentes aos estágios iniciais da doença com alterações sutis e pouco visíveis [Brejneboel et al. 2024, Beyaz et al. 2025], que são justamente os mais relevantes para intervenções preventivas precoces.

Com o avanço da Inteligência Artificial, métodos baseados em aprendizado profundo têm sido investigados para auxiliar especialistas na análise automatizada de radiografias de joelho [Touahema et al. 2024]. Arquiteturas como a EfficientNet combinadas a *transfer learning* (TL) têm demonstrado melhorias expressivas na classificação de imagens médicas [Diniz et al. 2024, Hassan and Ghadiri 2025]. Além disso, a extração automática de *Region of Interest* (ROI) é uma etapa essencial para direcionar o aprendizado às estruturas articulares relevantes, sendo que a classificação nos cinco graus da escala KL em raio-X ainda apresenta desafios como o desbalanceamento entre os graus e a semelhança visual entre classes adjacentes [Silvério and Machado 2025].

Para contornar o desafio da semelhança visual entre os graus da escala KL e destacar alterações articulares sutis, levanta-se a hipótese de que mecanismos de atenção podem ser incorporados às *Convolutional Neural Networks* (CNNs) para enfatizar características discriminantes e, conseqüentemente, melhorar a classificação dos graus intermediários da escala.

Diante desse contexto, o objetivo deste trabalho é desenvolver e avaliar um método automático para a classificação da OAJ nos cinco graus da escala KL em radiografias de joelho. O método proposto apresenta as seguintes contribuições:

- Integrar técnicas de pré-processamento de imagem com o objetivo de realçar as estruturas das radiografias de OAJ e melhorar a classificação das CNNs;
- Empregar EfficientNetV2B0 com módulo *Efficient Channel Attention* (ECA) e otimização bayesiana via Optuna, utilizando pesos de classe balanceados para melhorar a discriminação entre graus adjacentes da escala KL.

2. Trabalhos Relacionados

A classificação da OAJ em imagens de raio-X tem sido amplamente investigada por meio de CNNs. A seguir, são discutidos os principais trabalhos relacionados.

Nesse contexto, [Rani et al. 2024] exploraram uma CNN customizada de doze camadas projetada para realizar a predição multiclasse da severidade da doença. Na classificação global dos cinco graus da escala KL, o método alcançou uma acurácia de 78,40%, evidenciando o vasto potencial das arquiteturas profundas aplicadas a essas imagens. Em uma vertente semelhante, o estudo de [Lee et al. 2024a] desenvolveu um modelo combinando módulos independentes do tipo *plug-in* com as arquiteturas base SwinNet e EfficientNet. Os autores aplicaram vetores de suavização de rótulos visando atenuar ambigüidades clínicas, atingindo uma acurácia global de 75,70%.

Com foco em otimização, [Kumari et al. 2024] aplicaram *fine-tuning* das arquiteturas EfficientNet B5, B6 e B7. A pesquisa empregou técnicas de aumento de dados em

radiografias para a classificação multiclasse, com a versão B7 registrando o melhor desempenho, com acurácia de 78,53% e *F1-score* de 76,96%. Visando a integração clínica, [Lee et al. 2024b] validaram a viabilidade de plataformas *no-code* ao treinar a rede ResNet101. A abordagem alcançou acurácias variando entre 71,06% e 87,64% e especificidade de até 90,29% no conjunto de validação.

Em abordagem distinta, [Fei et al. 2024] avaliaram as redes VGG16, EfficientNetV2, ResNet34 e DenseNet196 para gerar uma pontuação contínua de severidade, permitindo um acompanhamento mais gradual. Nesse cenário, a EfficientNet obteve os resultados mais promissores, alcançando 71,00% de acurácia após o arredondamento das predições. Por outro lado, o desenvolvimento de um modelo leve foi o foco de [Abdusalomov et al. 2025], que aprimoraram a rede EfficientNet-B0 com um módulo de atenção de canais. Focado em atletas, o estudo formulou a detecção de osteoartrite inicial como uma tarefa de classificação binária, alcançando assim 91,50% de acurácia nessa delimitação específica.

[Momenpour and Abu Mallouh 2025] otimizaram o diagnóstico multiclasse utilizando a rede EfficientNet-B5 aliada ao *framework* de curadoria de dados CleanLab. A ferramenta corrigiu automaticamente inconsistências nos rótulos de uma base de dados pública (*Knee Osteoarthritis Dataset with Severity Grading*) contendo 9.786 radiografias de joelho. Ao mitigar os ruídos de anotação intrínsecos à subjetividade da escala clínica, o modelo otimizado alcançou acurácia de 82,07% e *F1-score* de 80,19%, superando de forma eficiente as abordagens concorrentes treinadas apenas com dados ruidosos.

A literatura atual de classificação da OAJ apresenta limitações recorrentes, como o desbalanceamento amostral e a aplicação isolada de técnicas de atenção e otimização. Além disso, a concentração de estudos em bases de larga escala deixa lacunas sobre o desempenho de modelos em conjuntos mais restritos e com múltiplas anotações de especialistas. Esse cenário restringe a compreensão sobre a generalização das arquiteturas em diferentes contextos de aquisição e protocolos de rotulagem.

Diante disso, o presente trabalho propõe um *pipeline* integrado que combina extração automática de ROI, *Contrast Limited Adaptive Histogram Equalization* (CLAHE) [Zuiderveld 1994], filtro de nitidez *Unsharp Masking* [Polesel et al. 2000] e normalização *Z-Score* [Jain et al. 1999] à arquitetura EfficientNetV2B0 [Tan and Le 2021] com módulo ECA [Wang et al. 2020]. A solução utiliza otimização Bayesiana via *framework* Optuna [Akiba et al. 2019] e treinamento final com pesos de classe balanceados, buscando maior equilíbrio entre as classes e melhor discriminação nos graus intermediários da escala KL.

3. Materiais e Método

Nesta seção, são apresentados os materiais e o método proposto, divididos em quatro etapas principais. Primeiramente, descreve-se a base de dados, em seguida, detalha-se o pré-processamento e a extração de ROI. Posteriormente, aborda-se a classificação com uma CNN integrada ao ECA e otimizada via Optuna. Por fim, calculam-se as métricas de validação. A Figura 1 ilustra estas etapas, que são detalhadas a seguir.

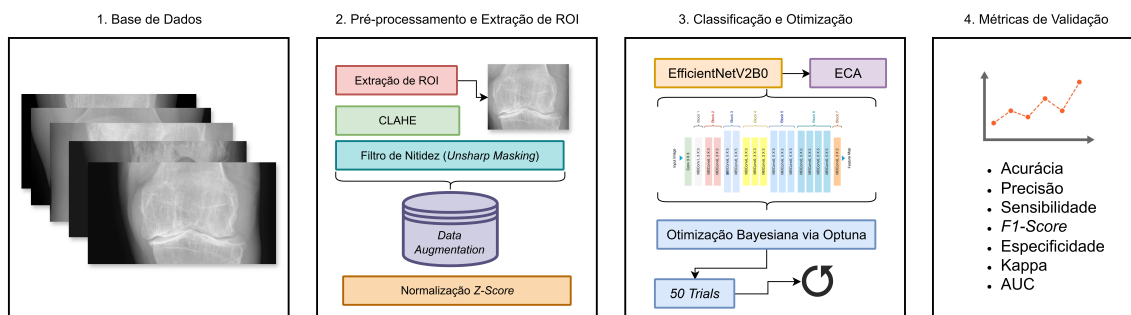


Figura 1. Ilustração do método proposto.

3.1. Base de Dados

Para a avaliação do método proposto, foi utilizada a base de dados *Digital Knee X-ray Images* [Gornale et al. 2020], composta por 1.650 imagens digitais de raio-X do joelho, coletadas em hospitais e centros de diagnóstico, anotadas de forma independente por dois especialistas ortopédicos. As anotações seguem a escala KL, que classifica a severidade da osteoartrite em cinco graus: Grau 0, Normal; Grau 1, Duvidoso (*Doubtful*); Grau 2, Leve (*Mild*); Grau 3, Moderado (*Moderate*); e Grau 4, Severo (*Severe*). As imagens originais possuem resoluções de 640 x 161 e 300 x 162 *pixels*. A Figura 2 apresenta exemplos das imagens na escala KL.

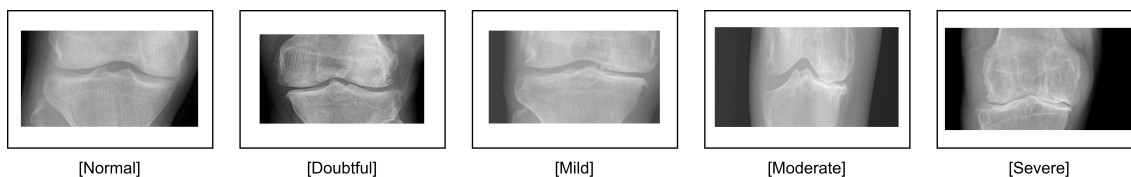


Figura 2. Exemplos de imagens da base *Digital Knee X-ray Image*.

Devido a pequenas divergências de avaliação nos graus iniciais, a base disponibiliza de forma independente as anotações nativas de cada especialista. Para a condução dos experimentos, os dados foram divididos de forma estratificada em 80% para treinamento e 20% para teste. Ressalta-se que, devido à ausência de identificadores de pacientes na base de dados, a separação foi realizada estritamente em nível de imagem, consistindo em uma limitação inerente à base utilizada. A Tabela 1 mostra a distribuição da base de dados por especialista.

Tabela 1. Quantidade de imagens segundo cada especialista

| Especialista | Quantidade de Imagens | | | | |
|----------------|-----------------------|--------|--------|--------|--------|
| | Grau 0 | Grau 1 | Grau 2 | Grau 3 | Grau 4 |
| Especialista 1 | 514 | 477 | 232 | 221 | 206 |
| Especialista 2 | 503 | 488 | 232 | 221 | 206 |

3.2. Pré-processamento e Extração de ROI

O pré-processamento consiste em remover áreas irrelevantes e realçar características morfológicas essenciais da imagem, como o espaço articular e a presença de osteófitos. Inicialmente, a imagem é convertida para tons de cinza e submetida a uma limiarização binária

de intensidade superior a 10 na escala de cinza. O maior contorno externo é identificado para traçar uma *bounding box*, permitindo o recorte da ROI e eliminando as bordas escuras.

O algoritmo CLAHE é aplicado, com limite de contraste de 2,0 e grade de 8×8 , com o intuito de atenuar variações de iluminação e melhorar o contraste local. Em seguida, as bordas ósseas são destacadas utilizando *Unsharp Masking*, que combina a imagem da etapa anterior com peso de valor 1,5 com uma versão suavizada por desfoque Gaussiano (*kernel* de 5×5 e *sigma* de 1,0, com peso de -0,5). Em seguida, aplica-se a normalização *Z-Score*, centralizando a distribuição dos *pixels* da imagem com média zero e desvio padrão um. Esta etapa garante a invariância ao contraste global e maior estabilidade numérica durante a descida do gradiente no treinamento da rede.

Para o treino, aplicam-se técnicas de *Data Augmentation*, visando aumentar a variabilidade das amostras e prevenir *overfitting*. As transformações incluem espelhamento horizontal, rotações aleatórias com fator de 0,05, além de translações e *zoom* de até 10%. A Figura 3 ilustra as técnicas aplicadas.

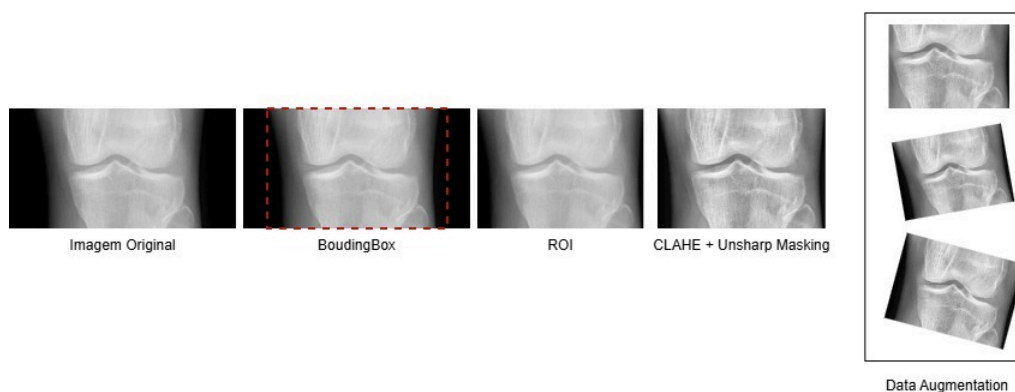


Figura 3. Ilustração da etapa de Pré-processamento e Extração de ROI.

3.3. Classificação e Otimização

Nesta seção, descreve-se o processo de classificação das imagens e a otimização do modelo. Primeiramente, apresentam-se a arquitetura EfficientNetV2B0 avaliada, seguido da estratégia de otimização de hiperparâmetros e, por fim, o protocolo de validação adotado.

3.3.1. Extração de Características

A abordagem proposta emprega a EfficientNetV2B0 como extratora de características, inicializada com pesos pré-treinados no conjunto de dados ImageNet [Deng et al. 2009] por meio da técnica de TL. Para aprimorar a representatividade dos mapas de características gerados, integrou-se o ECA logo após a saída do *backbone*.

Diferentemente de outros mecanismos de atenção convencionais, o ECA atua capturando a interação cruzada local entre canais sem realizar redução de dimensionalidade. Isso permite que a rede enfatize padrões relevantes, como o estreitamento articular e a presença de osteófitos. Em seguida, os mapas atencionais são repassados para a cabeça

classificadora (*Classifier head*), que é composta por uma camada de agrupamento global médio (*Global Average Pooling*), uma camada de regularização *Dropout* e uma camada densa final com função de ativação *Softmax*, responsável por mapear as características processadas para as probabilidades das cinco categorias da escala KL. Na fase inicial do treinamento, os parâmetros das camadas inferiores do *backbone* são congelados, atualizando apenas os pesos referentes ao módulo de atenção e à classificação.

Durante o treinamento, também é levado em conta o desbalanceamento da base de dados. O desbalanceamento amostral entre os diferentes graus da escala KL é tratado diretamente na função de perda, através do cálculo automático de pesos de classe balanceados (*balanced class weights*), o que atribui maior penalidade aos erros preditivos nas classes minoritárias.

3.3.2. Otimização de Hiperparâmetros

Em uma arquitetura CNN, a configuração dos hiperparâmetros influencia diretamente a sua generalização e a eficácia no aprendizado de características. Para mitigar as limitações do ajuste manual, utiliza-se o *framework* Optuna [Akiba et al. 2019] para realizar a busca de hiperparâmetros por meio da Otimização Bayesiana. Este método explora um espaço de busca de maneira probabilística, avaliando o histórico das tentativas anteriores para sugerir combinações cada vez mais favoráveis.

Durante o processo, configurou-se um total de 50 *trials* com o intuito de maximizar a métrica de acurácia do modelo. O espaço de busca avaliado incluiu: *learning rate* (LR), tamanho de lote (*batch size*), taxa de regularização (*dropout*), decaimento de peso (*weight decay*) e a escolha do algoritmo otimizador (Adam, AdamW e *Stochastic Gradient Descent*). Após a identificação dos melhores hiperparâmetros, o treinamento final foi realizado, com ajuste dinâmico de LR via *ReduceLROnPlateau* e *Early Stopping* monitorando a acurácia de validação.

3.4. Métricas de Validação

Para avaliar o desempenho do método proposto, foram utilizadas métricas de validação consolidadas na literatura para problemas de classificação diagnóstica. São elas: acurácia (ACC), precisão (PRE), sensibilidade (SEN) ou *recall*, *F1-Score*, especificidade (ESP), Kappa (KAP) e a área sob a curva ROC (AUC).

4. Resultados e Discussão

Nesta seção, são descritos os experimentos realizados para validar o método proposto, apresentando as métricas de validação obtidas na arquitetura avaliada. Adicionalmente, são realizadas comparações com trabalhos relacionados e estudos de casos.

4.1. Ambiente de Treinamento

Os experimentos foram executados no ambiente de computação em nuvem Google Colab, utilizando GPU Nvidia A100. A implementação foi realizada em linguagem Python, com uso das bibliotecas TensorFlow, Keras, scikit-learn, OpenCV, Pillow, NumPy, Matplotlib e Optuna.

4.2. Hiperparâmetros e Regularização

Para o *baseline*, os hiperparâmetros foram fixados manualmente, adotando valores padrões utilizados na literatura: LR de 1×10^{-2} , *batch size* de 32, taxa de *dropout* de 0,5, *weight decay* de 1×10^{-3} e otimizador AdamW.

Para o método proposto, o Optuna realizou 50 *trials*, maximizando a acurácia média de validação. A Tabela 2 apresenta os melhores hiperparâmetros encontrados referentes a cada especialista.

Tabela 2. Melhores hiperparâmetros encontrados pelo Optuna.

| Especialista | LR | Batch Size | Weight Decay | Dropout | Otim. |
|----------------|-----------------------|------------|-----------------------|---------|-------|
| Especialista 1 | $7,00 \times 10^{-4}$ | 64 | $1,00 \times 10^{-4}$ | 0,13 | Adam |
| Especialista 2 | $9,50 \times 10^{-3}$ | 64 | $4,00 \times 10^{-4}$ | 0,11 | Adam |

Nota-se que a Otimização Bayesiana convergiu para LR's menores em relação ao *baseline*, sugerindo que a EfficientNetV2B0 exige um ajuste mais fino dos pesos quando combinada ao ECA e às técnicas de pré-processamento. Além disso, o Optuna selecionou o otimizador Adam e taxas de *dropout* mais baixas, indicando que a arquitetura não necessita de regularizações tão severas nesse cenário.

4.3. Resultados Experimentais

Para validar o método proposto, foram conduzidos dois experimentos. No primeiro, avalia-se o desempenho da EfficientNetV2B0 sem as contribuições propostas em seu estado original. No segundo, avalia-se o método proposto, integrando a extração de ROI, CLAHE, *Unsharp Masking*, ECA e Otimização Bayesiana de hiperparâmetros via Optuna. Em ambas as abordagens, adotou-se a função de perda *Categorical Crossentropy*, adequada para a classificação multiclasse nos cinco graus da escala KL.

4.3.1. Resultados do Baseline

A Tabela 3 apresenta os resultados do *baseline*, sem a aplicação de extração de ROI, CLAHE, *Unsharp Masking*, ECA e Otimização Bayesiana de hiperparâmetros via Optuna.

Tabela 3. Resultados do *baseline*.

| Especialista | ACC | PRE | SEN | F1-Score | ESP | KAP | AUC |
|----------------|--------|--------|--------|----------|--------|--------|--------|
| Especialista 1 | 61,52% | 62,79% | 61,52% | 61,83% | 89,81% | 49,53% | 86,85% |
| Especialista 2 | 63,64% | 63,27% | 63,64% | 61,64% | 90,25% | 52,09% | 88,44% |

As métricas apontam que, na ausência das contribuições propostas, a EfficientNetV2B0 atinge um desempenho mediano e equilibrado entre os avaliadores. O treinamento baseado nas anotações do Especialista 2 alcançou ligeira vantagem em acurácia, Kappa e AUC. O resultado global inferior, com destaque para o coeficiente Kappa (na faixa de 50%), comprova a limitação do modelo base em diferenciar as classes da escala KL sem a integração do pré-processamento, do módulo ECA e do ajuste via Optuna.

4.3.2. Resultados do Método Proposto

A Tabela 4 apresenta os resultados obtidos pelo método proposto, integrando extração de ROI, CLAHE, *Unsharp Masking*, ECA e Otimização Bayesiana de hiperparâmetros via Optuna.

Tabela 4. Resultados do método proposto.

| Especialista | ACC | PRE | SEN | F1-Score | ESP | KAP | AUC |
|----------------|--------|--------|--------|----------|--------|--------|--------|
| Especialista 1 | 73,33% | 73,69% | 73,33% | 73,46% | 92,95% | 65,08% | 92,18% |
| Especialista 2 | 77,58% | 77,82% | 77,58% | 76,98% | 94,00% | 70,48% | 93,99% |

Os resultados revelam que o método proposto obteve ganhos expressivos no desempenho da EfficientNetV2B0. A avaliação baseada no Especialista 2 obteve o melhor cenário geral, atingindo 77,58% de acurácia, 76,98% de *F1-Score* e 70,48% de Kappa, o que representa incrementos de aproximadamente 14 e 15 pontos em relação ao *baseline*. Para o Especialista 1, o avanço também foi notável, com o coeficiente Kappa saltando de 49,53% para 65,08%. Essa evolução comprova que a união entre o pré-processamento, a atenção por canal e a otimização de hiperparâmetros aumentou a capacidade discriminativa do modelo diante das variações da escala KL.

As Figuras 4 e 5 apresentam as matrizes de confusão normalizadas para ambos os especialistas. Observa-se que as previsões incorretas concentram-se majoritariamente em classes adjacentes, como, por exemplo, o Grau 1 predito como Grau 0 ou 2, evidenciando que os erros do modelo ocorrem em estágios de transição da escala KL, onde a ambiguidade clínica e a subjetividade visual entre os especialistas são mais acentuadas.

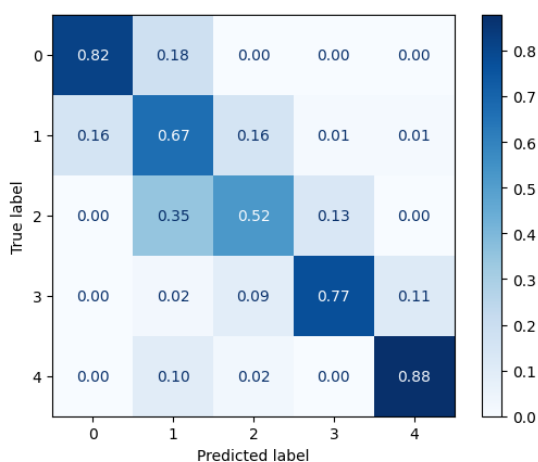


Figura 4. Matriz de confusão - Especialista 1.

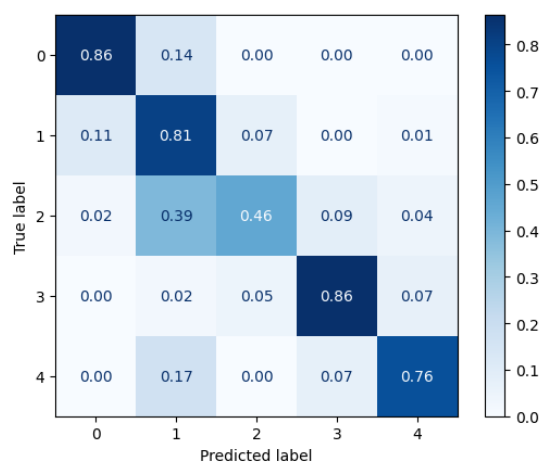


Figura 5. Matriz de confusão - Especialista 2.

A menor predominância de erros em níveis distantes confirma a estabilidade da rede. O maior ponto de confusão reside na fronteira entre os Graus 1 e 2 (chegando a 39% no Especialista 2), o que é clinicamente esperado, dado que esses estágios compartilham características radiológicas limítrofes e de difícil distinção na escala KL.

4.4. Comparação com a Literatura

Como discutido na Seção 2, diversos estudos investigam a classificação de OAJ utilizando bases de dados distintas, tais como *Osteoarthritis Initiative* (OAI), *Multicenter Osteoarthritis Study* (MOST), *Knee Osteoarthritis Dataset* (KOD) e *Knee Osteoarthritis Severity Grading* (KOSG). A Tabela 5 resume as metodologias adotadas e os desempenhos reportados na literatura.

Tabela 5. Comparação do método proposto com os trabalhos relacionados.

| Trabalho | Base de Dados | Técnica | Divisão | ACC |
|----------------------------------|---------------------------|--|--------------------|---------------|
| [Rani et al. 2024] | OAI | CNN Customizada (12 Camadas) | Multiclasse | 78,40% |
| [Rani et al. 2024] | OAI | CNN Customizada (12 Camadas) | Binário | 92,30% |
| [Lee et al. 2024a] | OAI / MOST | <i>Ensemble</i> (Swin + EfficientNet) | Multiclasse | 75,70% |
| [Kumari et al. 2024] | KOD | EfficientNet-B7 + TL | Multiclasse | 78,53% |
| [Lee et al. 2024b] | OAI | ResNet101 (<i>No-code</i>) | Multiclasse | 71,06% |
| [Fei et al. 2024] | OAI | EfficientNetV2 + Regressão | Regressão | 71,00% |
| [Abdusalomov et al. 2025] | KOSG | EfficientNet-B0 + ECA | Binário | 91,50% |
| [Momenpour and Abu Mallouh 2025] | KOD | EfficientNet-B5 + CleanLab | Multiclasse | 82,07% |
| Proposto (Esp. 1) | Digital Knee X-ray | EfficientNetV2B0 + ECA + Optuna | Multiclasse | 73,33% |
| Proposto (Esp. 2) | Digital Knee X-ray | EfficientNetV2B0 + ECA + Optuna | Multiclasse | 77,58% |

A análise da Tabela 5 evidencia que estudos com altas taxas de acurácia, como [Rani et al. 2024] e [Abdusalomov et al. 2025], simplificam o problema para uma classificação binária, o que eleva substancialmente as métricas de desempenho. Além disso, parte das abordagens focadas no cenário multiclasse negligencia o pré-processamento direcionado, como a extração de ROI e o realce das estruturas articulares, o que dificulta a distinção correta dos graus intermediários da escala KL.

Embora trabalhos como [Momenpour and Abu Mallouh 2025] e [Kumari et al. 2024] reportem acurácias superiores na classificação dos cinco graus KL, eles dependem de arquiteturas com elevada complexidade computacional e se beneficiam de bases de dados substancialmente maiores, as quais não disponibilizam anotações duplas independentes. Por esse motivo, tais estudos não especificam a vertente de especialistas utilizada para os rótulos de referência. Em contrapartida, o método proposto atinge até 77,58% de acurácia em um conjunto de imagens mais restrito e com dupla avaliação de especialistas, empregando uma rede notavelmente mais leve. Essa eficiência, combinada ao módulo ECA e ao ajuste via Optuna, consolida o modelo como uma alternativa viável e competitiva nesse cenário.

4.5. Estudo de Casos

Nesta seção, é apresentado o estudo de caso com as predições do método proposto. A Figura 6 ilustra exemplos de classificações corretas e incorretas da severidade da osteoartrite nos graus da escala KL.

Tanto em (A) quanto em (B), o modelo classificou corretamente as imagens pertencentes aos graus mais consolidados da OAJ. Esse comportamento demonstra que a rede é capaz de identificar adequadamente padrões morfológicos claros, como o espaço articular preservado ou o desgaste severo com presença de osteófitos.

Em contrapartida, os casos de erro à direita em (A) e (B) envolvem o Grau 1 (Dúvidoso), predito como Grau 0 e Grau 2. Essa confusão ocorre porque a transição entre esses níveis é marcada pelo surgimento precoce de osteófitos e um estreitamento sutil do

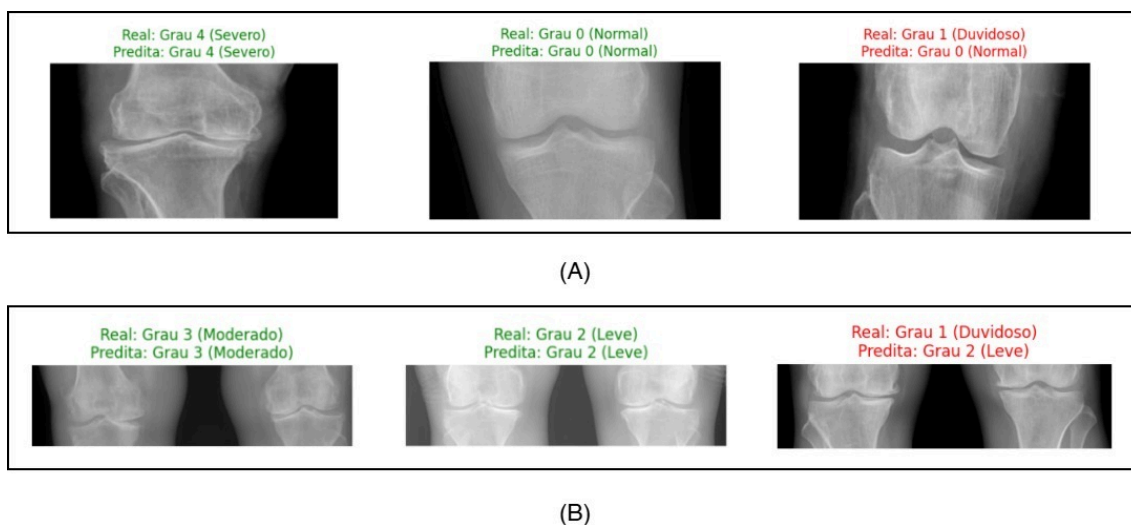


Figura 6. (A) Inferências de um único joelho e (B) Inferências de ambos os joelhos.

espaço articular, sinais que podem ser interpretados de forma divergente conforme a subjetividade clínica ou a nitidez da imagem. Sob uma perspectiva clínica, a concentração de erros em classes adjacentes sugere que o modelo apresenta um comportamento coerente com as incertezas diagnósticas inerentes à escala KL, preservando a segurança da triagem ao evitar falhas críticas em estágios de maior severidade.

5. Conclusão

Este trabalho apresentou um método automático para a classificação de severidade da OAJ na base de dados *Digital Knee X-ray Images*, integrando extração de ROI, técnicas de realce (CLAHE e *Unsharp Masking*), arquitetura EfficientNetV2B0 com módulo ECA e Otimização Bayesiana via Optuna. Os resultados comprovam a eficácia da abordagem, alcançando acurácia de 77,58%, *F1-Score* de 76,98% e Kappa de 70,48% na avaliação do Especialista 2. Esses valores representam ganhos expressivos em relação ao *baseline* e mantêm um desempenho competitivo frente à literatura, equilibrando precisão e baixo custo computacional. Isso evidencia que a união entre o pré-processamento e a atenção por canal aumenta a capacidade do modelo em distinguir as alterações sutis inerentes à escala KL.

Como trabalhos futuros, propõe-se a investigação de abordagens utilizando *ensemble* e a aplicação de métodos de explicabilidade visual para mapear as regiões de incerteza da rede, especialmente nos graus iniciais da doença. Além disso, propõe-se a avaliação do método proposto em bases de dados de maior escala, visando atestar sua capacidade de generalização e ampliar sua viabilidade como ferramenta de suporte ao diagnóstico clínico.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, Fundação de Amparo a Pesquisa do Maranhão (FAPEMA), Conselho Nacional de Desenvolvimento

Científico e Tecnológico (CNPq). Ainda, reconhecemos o uso do LLM para revisão ortográfica, correção gramatical e assistência na tradução de termos específicos.

Referências

- Abdusalomov, A., Mirzakhilov, S., Umirzakova, S., Ismailov, O., Sultanov, D., Nasimov, R., and Cho, Y. I. (2025). Lightweight early detection of knee osteoarthritis in athletes. *Scientific Reports*, 15(1):31413.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2623–2631.
- Beyaz, S., Yayli, S. B., and Kılıç, K. (2025). From variability to consistency: building a kellgren-lawrence gonarthrosis dataset. *Journal of Orthopaedic Surgery and Research*, 20(1):1–9.
- Brejnebo, M. W., Lenskjold, A., Ziegeler, K., Ruitenbeek, H., Müller, F. C., Nybing, J. U., Visser, J. J., Schiphouwer, L. M., Jasper, J., Bashian, B., et al. (2024). Interobserver agreement and performance of concurrent ai assistance for radiographic evaluation of knee osteoarthritis. *Radiology*, 312(1):e233341.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.
- Diniz, J. O., Ribeiro, N. P., Junior, D. A. D., da Cruz, L. B., de Carvalho Filho, A. O., Gomes Jr, D. L., Silva, A. C., and de Paiva, A. C. (2024). Efficientxyz-deepfeatures: seleção de esquema de cor e arquitetura deep features na classificação de câncer de cólon em imagens histopatológicas. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 82–93. SBC.
- Fei, M., Lu, S., Chung, J. H., Hassan, S., Elsisy, J., and Schneiderman, B. A. (2024). Diagnosing the severity of knee osteoarthritis using regression scores from artificial intelligence convolution neural networks. *Orthopedics*, 47(5):e247–e254.
- Gornale, S. S., Patravali, P. U., and Hiremath, P. S. (2020). A comprehensive digital knee x-ray image dataset for the assessment of osteoarthritis. *JSM Biomed. Imag. Data Pap*, 6:1012.
- Hassan, E. and Ghadiri, H. (2025). Advancing brain tumor classification: A robust framework using efficientnetv2 transfer learning and statistical analysis. *Computers in Biology and Medicine*, 185:109542.
- Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323.
- Kellgren, J. and Lawrence, J. (1957). Radiological assessment of rheumatoid arthritis. *Annals of the rheumatic diseases*, 16(4):485.
- Kumari, L. V. R., Jagruti, K., Chandra, G. R., Reddy, M. S., and Bhadramma, B. (2024). Transfer learning based efficientnet for knee osteoarthritis classification. *Traitement du Signal*, 41(2).

- Langworthy, M., Dasa, V., and Spitzer, A. I. (2024). Knee osteoarthritis: disease burden, available treatments, and emerging options. *Therapeutic Advances in Musculoskeletal Disease*, 16:1759720X241273009.
- Lee, D. W., Song, D. S., Han, H.-S., and Ro, D. H. (2024a). Accurate, automated classification of radiographic knee osteoarthritis severity using a novel method of deep learning: Plug-in modules. *Knee Surgery & Related Research*, 36(1):24.
- Lee, K.-H., Lee, R.-W., Yun, J.-S., Kim, M.-S., and Choi, H.-S. (2024b). Automated diagnosis of knee osteoarthritis using resnet101 on a deep: Phi: leveraging a no-code ai platform for efficient and accurate medical image analysis. *Diagnostics*, 14(21):2451.
- Li, E., Tan, J., Xu, K., Pan, Y., and Xu, P. (2024). Global burden and socioeconomic impact of knee osteoarthritis: a comprehensive analysis. *Frontiers in Medicine*, 11:1323091.
- Momenpour, T. and Abu Mallouh, A. (2025). Optimizing cnn-based diagnosis of knee osteoarthritis: Enhancing model accuracy with cleanlab relabeling. *Diagnostics*, 15(11):1332.
- Polesel, A., Ramponi, G., and Mathews, V. J. (2000). Image enhancement via adaptive unsharp masking. *IEEE transactions on image processing*, 9(3):505–510.
- Rani, S., Memoria, M., Almogren, A., Bharany, S., Joshi, K., Altameem, A., Rehman, A. U., and Hamam, H. (2024). Deep learning to combat knee osteoarthritis and severity assessment by using cnn-based classification. *BMC Musculoskeletal Disorders*, 25(1):817.
- Silvério, A. C. M. and Machado, A. M. C. (2025). Classificação da osteoartrite de joelho em imagens de raio-x por meio de ensemble learning. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 973–978. SBC.
- Tan, M. and Le, Q. (2021). Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR.
- Touahema, S., Zaimi, I., Zrira, N., Ngote, M. N., Douhousne, H., and Aouial, M. (2024). Medknee: A new deep learning-based software for automated prediction of radiographic knee osteoarthritis. *Diagnostics*, 14(10):993.
- Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11534–11542.
- Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc.