Cost-Effective Screening for Knee Osteoarthritis: A Serial Clinical-Image AI Pipeline

Júlio Guerra Domingues
Department of Computer Science
Department of Anatomy and Medical Imaging
UFMG
Belo Horizonte, Brazil
juliogdomingues@gmail.com

Rosa Weiss Telles
Department of Clinical Medicine
UFMG
Belo Horizonte, Brazil
rwtelles1974@gmail.com

Sandhi Maria Barreto
Department of Preventive and Social Medicine
UFMG
Belo Horizonte, Brazil
sandhi.barreto@gmail.com

Adriano Alonso Veloso
Department of Computer Science
UFMG
Belo Horizonte, Brazil
adrianov@dcc.ufmg.br

***Abstract*—Knee osteoarthritis (KOA) is a high-impact degenerative disease whose diagnosis can be improved with artificial intelligence (AI). Existing models, however, often lack validation for the Brazilian population and efficient integration into clinical workflows. This study presents the development and validation of a multimodal model for KOA diagnosis using radiographic images and clinical-epidemiological data from the ELSA-Brasil MSK study. The methodology employed a serial triage architecture: a clinical model (Logistic Regression optimized via Minimal Predictive Model Search [MPMS]) acts as a high-sensitivity filter, followed by a Deep Learning image model (DenseNet-161) for final confirmation. The image model achieved an Area Under the ROC Curve (AUC-ROC) of 0.866, while the optimized clinical model achieved an AUC-ROC of 0.827. In a simulated triage scenario, this serial approach reduced the need for radiographic exams by 44.5% while maintaining a high overall sensitivity of 84.5% and achieving a final specificity of 99.7%. These results demonstrate that a multimodal AI approach can provide a robust, cost-effective, and validated tool for KOA screening in the Brazilian context.**

***Index Terms*—Knee Osteoarthritis, Deep Learning, Multimodal, Artificial Intelligence, ELSA-Brasil, Serial Triage.**

I. Introduction

Osteoarthritis (OA) is one of the most prevalent musculoskeletal diseases worldwide, with a major impact on quality of life, especially in the elderly [1]. Knee involvement (KOA) is notable for its high frequency and associated morbidity, being one of the leading causes of physical disability globally [2]. Symptoms such as pain, stiffness, and loss of mobility lead to significant economic and social repercussions, including treatment costs and loss of productivity [3]–[5].

The diagnosis of KOA is traditionally performed by correlating clinical findings with the analysis of imaging exams, such as radiographs, which require interpretation by medical specialists [6], [7]. In large-scale longitudinal studies, such as the Longitudinal Study of Adult Health (ELSA-Brasil), this process can become a bottleneck, being time-consuming, costly, and subject to inter-rater variability [8].

The application of artificial intelligence (AI), particularly deep learning techniques like Convolutional Neural Networks (CNNs), has shown promise in the classification of medical images [9]. In a previous study, our team developed a CNN model for classifying knee radiographs in the context of the ELSA-Brasil MSK study, with satisfactory results [8]. However, most existing models in the literature have been trained and validated on North American and European populations, which limits their generalizability to other demographic contexts, such as Brazil's [9], [10].

This work presents a significant advancement over the previous approach by developing a multimodal pipeline that integrates image data with clinical-epidemiological information. Unlike simple fusion, we propose a serial triage workflow to optimize resources. The objective is to create a diagnostic tool with higher accuracy and efficiency, providing a validated model for the Brazilian population using the rich database of the ELSA-Brasil MSK.

II. Methods

A. Study Design and Sample

This study utilizes radiographs from the first wave (2012-2014) and clinical-epidemiological data from both the first and second waves (2017-2019) of the ELSA-Brasil Musculoskeletal (ELSA-Brasil MSK) study.

The baseline cohort (first wave) included 2901 active and retired civil servants, 53% women, with ages ranging from 38 to 79 years (mean: 56.0, SD: 8.9). From this baseline, the present study uses a dataset of 5660 knee radiographs obtained from 2830 participants. Clinical follow-up data from the second wave were collected from 2618 of these participants.

The radiographic readings from the first wave have been completed, with radiographic KOA classified according to international calibration and classification standards [11].

B. Ethical Aspects

The ethical principles of respect for persons, beneficence, and justice were considered in the planning of both ELSA-Brasil and ELSA-Brasil MSK. ELSA-Brasil was approved by the Research Ethics Committees of the six participating teaching and research institutions. The baseline of the ELSA-Brasil MSK ancillary study and its first follow-up were approved through an amendment submitted to the Research Ethics Committee (REC) of UFMG (amendment CAAE 0186.1.203.000-06 of proposal ETIC 186/06, dated 09/03/2012, and CAAE 47125015.4.1001.5149, opinion 1.897.023 of 01/24/2017).

C. Data Collection and Interpretation

1) Clinical and Epidemiological Data: The assessments conducted in the cohort include sociodemographic characteristics, health-related behaviors, psychosocial data, cognitive function, anthropometric measurements, medications, subclinical and laboratory measurements, and the incidence of non-communicable diseases. The ELSA-Brasil MSK adds important assessments of musculoskeletal disorders and potential risk factors. Participants were evaluated for prior medical diagnoses of musculoskeletal diseases, history of pain, discomfort, and/or stiffness.

2) Radiographic Examination and Interpretation: The first-wave radiographs used in this study were acquired in a bilateral posteroanterior fixed-flexion view. The interpretation was performed according to a validated protocol: (i) screening for "possible OA" by two technologists; and (ii) review by a radiologist. The inter-observer kappa was 0.755 (95% CI 0.663-0.847) [12].

D. Computational Model

The proposed computational pipeline integrates two data sources (radiographic images and tabular clinical-epidemiological data) into a serial multimodal framework.

1) Image Branch: The image processing component of our pipeline leverages a previously validated deep learning model developed by our team [8]. This model is based on the DenseNet-161 architecture [13], which was fine-tuned using knee radiographs from the ELSA-Brasil MSK cohort. The detailed methodology, including pre-processing steps (knee individualization, Region of Interest [ROI] extraction, and data augmentation) and performance metrics, is fully described in our prior publication [8]. In the current work, we integrate this pre-trained model as the confirmatory stage of our serial triage workflow.

2) Clinical/Tabular Data Branch: Variables were selected based on a literature review (see Appendix, Table II). The development of the clinical model followed a rigorous pipeline. First, missing clinical data were imputed using the median, while missing Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) scores were imputed to 0, assuming asymptomatic knees in the absence of pain reports. Feature selection involved a two-step process using LASSO (L1-penalized Logistic Regression) for initial screening followed by Minimal Predictive Model Search (MPMS) to optimize the trade-off between parsimony and performance. While MPMS identified a larger candidate set, we prioritized a parsimonious 5-variable Logistic Regression model for the final pipeline to ensure ease of application in clinical settings. This final classifier was trained and evaluated using Nested Cross-Validation (outer 5-fold, inner 3-fold) to ensure robust performance estimation.

3) Serial Triage Fusion: Instead of simple concatenation, a serial triage simulation was performed. The optimized clinical model acts as an initial filter with high sensitivity ($\approx$ 90%). Only participants flagged as "positive" by the clinical model are forwarded to the image model (CNN). This replicates a realistic clinical workflow where radiographs are ordered only when clinically indicated.

E. Performance Evaluation

Evaluation metrics included accuracy, sensitivity, specificity, and the area under the ROC curve (AUC). SHAP (SHapley Additive exPlanations) values were computed to interpret the clinical model's decisions.

III. Results

A. Unimodal Image Model

As detailed in our previous work [8], the DenseNet-161 model demonstrated high performance for diagnosing radiographic KOA (Kellgren-Lawrence [KL] grade ≥ 2). The model achieved an AUC-ROC of 0.866, with an accuracy of 90.7%, sensitivity of 93.8%, and specificity of 99.4% at the optimal operating point.

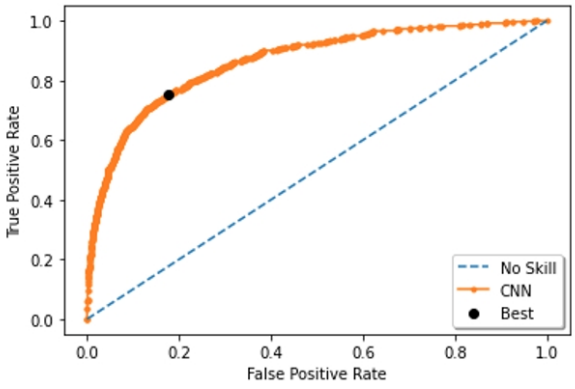

Fig. 1. ROC Curve of the Image Model (DenseNet-161) [8].

B. Clinical Model and Feature Selection

The feature selection process identified a core set of predictive variables. The MPMS algorithm selected an optimal model (MPMS-Best) containing 10 variables, achieving an AUC-ROC of 0.827. To favor interpretability, a simpler model with 5 variables was also evaluated, achieving a competitive

AUC-ROC of 0.818 (Fig. 2). The five core predictors identified were age (continuous), Body Mass Index (BMI), frequent knee pain, history of knee surgery, and history of knee trauma.

To facilitate clinical application and external validation, the source code and a web-based risk calculator are publicly available at https://github.com/juliogdomingues/kneelsa-clinical. The logistic regression equation parameters for this parsimonious model are detailed in Table I.

TABLE I
LOGISTIC REGRESSION MODEL COEFFICIENTS (5-VARIABLE MODEL).

Variable	Coefficient (β)	Odds Ratio (95% CI)
Intercept	-10.871	-
Age (years)	0.087	1.09 (1.08-1.10)
BMI (kg/m^2)	0.120	1.13 (1.10-1.15)
Frequent Knee Pain	0.363	1.44 (1.32-1.57)
History of Surgery	2.041	7.69 (5.14-11.52)
History of Trauma	0.767	2.15 (1.69-2.74)

Note: Probability $P = 1/(1 + e^{-(\beta_0 + \sum \beta_i X_i)})$

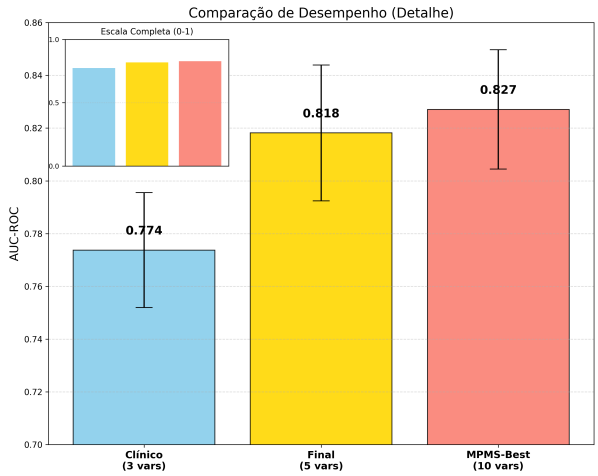

Fig. 2. Performance comparison of clinical models (AUC-ROC with 95% CI).

C. Interpretability (SHAP)

SHAP analysis confirmed that Age and BMI were the most dominant predictors, followed by frequent pain and history of mechanical injury (Fig. 3). Metabolic syndrome markers (Joint Interim Statement [JIS] criteria) showed contribution but with lower impact than mechanical and demographic factors.

D. Serial Triage Simulation

The simulation of the serial workflow (Clinical Model $\rightarrow$ Image Model) on a cohort of 1,000 subjects yielded significant efficiency gains. We adopted a sensitivity threshold of 90% for the clinical screening step, prioritizing the detection of true cases.

Regarding efficiency, the pipeline reduced the need for radiographic exams by 44.5%. In terms of performance, the final sensitivity of the combined system was 84.5%, with a minimal number of false negatives. Furthermore, the system

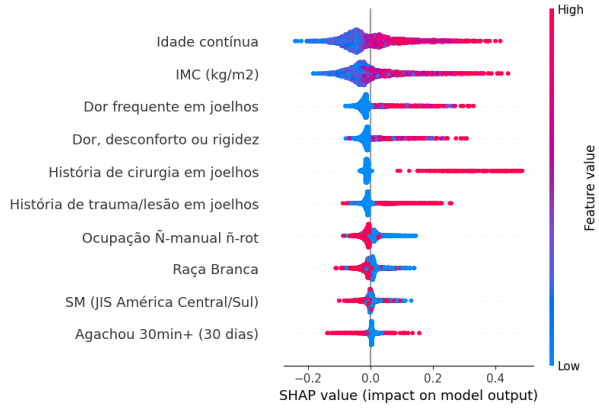

Fig. 3. SHAP Summary Plot showing feature importance for the clinical model.

demonstrated high safety and precision, achieving a final specificity of 99.7% and a Positive Predictive Value (PPV) of 97.7%, indicating a very low rate of false alarms.

IV. DISCUSSION

The evolution from a unimodal image model to a multimodal serial approach addresses a critical need in public health: optimizing resource allocation. While the image model provides structural precision (AUC 0.866), it is resource-intensive. The clinical model developed in this study proved to be a robust screening tool (AUC 0.827), capable of filtering out nearly half of the population who do not require radiation exposure or specialized image interpretation.

The choice of a 90% sensitivity threshold for the clinical screening represents a strategic trade-off. Our simulations showed that increasing sensitivity to 95% or 99% would drastically reduce the efficiency gain (exam reduction dropping to 32% and 14%, respectively), while a lower threshold (e.g., 80%) would result in an unacceptable number of missed cases (false negatives). The 90% point offers a balanced "sweet spot" for mass screening, but this parameter can be adjusted according to specific public health priorities and resource availability.

The selection of variables confirmed established risk factors. Age and BMI remain the strongest predictors, consistent with the literature [14], [15]. This biological plausibility was reinforced by the SHAP analysis, which confirmed that the model's reliance on these factors aligns with clinical intuition. Although "History of Surgery" presents the highest individual effect size (OR 7.69), Age and BMI showed higher global importance (SHAP) due to their ubiquitous presence and cumulative effect across the population range. Interestingly, while metabolic factors were included based on recent evidence of "metabolic OA" [16], the data-driven MPMS selection prioritized mechanical and historical factors (trauma/surgery) for the minimal model, suggesting these have higher immediate discriminative power in this specific cohort.

A. Limitations

As it is a cohort of civil servants, the ELSA-Brasil sample does not represent the entire Brazilian population, as the extremes of the social hierarchy are not represented. However, a previous study showed that the prevalence of risk factors for chronic non-communicable diseases in the cohort is similar to that found in the general adult population [17]. Furthermore, representativeness is not a requirement for making valid inferences in longitudinal studies [18].

V. Conclusions

This project successfully developed and validated a multimodal AI framework for KOA diagnosis. The integration of a parsimonious clinical model with a deep learning image model in a serial triage workflow resulted in a 44.5% reduction in unnecessary exams without significantly compromising diagnostic sensitivity. This approach offers a scalable, cost-effective solution for KOA screening in the Brazilian health system. We openly provide the clinical model for public validation to encourage its testing and refinement in diverse populations.

In future studies, we intend to fully integrate these models into a unified multimodal architecture capable of simultaneous processing. This evolution will allow the system to adapt dynamically to the clinical context, providing robust predictions regardless of whether radiographic imaging is immediately available.

References

[1] L. Santo and T. Okeyode, "National ambulatory medical care survey: 2018 national summary tables," [Internet], 2018. [Online]. Available: https://www.cdc.gov/nchs/data/ahcd/namcs_summary/2018-namcs-web-tables-508.pdf

[2] T. Vos, A. D. Flaxman, M. Naghavi, R. Lozano, C. Michaud, M. Ezzati, K. Shibuya, J. A. Salomon *et al.*, "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the global burden of disease study 2010," *Lancet*, vol. 380, no. 9859, pp. 2163–2196, dec 2012, erratum in: Lancet. 2013 Feb 23;381(9867):628.

[3] L. Sharma, "Osteoarthritis of the knee," *New England Journal of Medicine*, vol. 384, no. 1, pp. 51–59, jan 2021.

[4] D. J. Hunter, D. Schofield, and E. Callander, "The individual and socioeconomic impact of osteoarthritis," *Nature Reviews Rheumatology*, vol. 10, no. 7, pp. 437–441, jul 2014.

[5] L. M. March and C. J. Bachmeier, "Economics of osteoarthritis: a global perspective," *Baillieres Clinical Rheumatology*, vol. 11, no. 4, pp. 817–834, nov 1997.

[6] R. C. C. Miguel, L. A. Machado, L. Costa-Silva, R. W. Telles, and S. M. Barreto, "Performance of distinct knee osteoarthritis classification criteria in the ELSA-Brasil musculoskeletal study," *Clinical Rheumatology*, vol. 38, no. 3, pp. 793–802, mar 2019.

[7] C. Buckland-Wright, "Which radiographic techniques should we use for research and clinical practice?" *Best Practice & Research Clinical Rheumatology*, vol. 20, no. 1, pp. 39–55, feb 2006.

[8] J. G. Domingues, D. C. Araujo, L. Costa-Silva, A. M. C. Machado, L. A. C. Machado, A. A. Veloso, S. M. Barreto, and R. W. Telles, "Development of a convolutional neural network for diagnosing osteoarthritis, trained with knee radiographs from the ELSA-Brasil musculoskeletal," *Radiologia Brasileira*, vol. 56, no. 5, sep 2023.

[9] M. Binvignat, V. Pedoia, A. J. Butte, K. Louati, D. Klatzmann, F. Berenbaum, E. Mariotti-Ferrandiz, and J. Sellam, "Use of machine learning in osteoarthritis research: a systematic literature review," *RMD Open*, vol. 8, no. 1, p. e001998, mar 2022.

[10] A. Jamshidi, J. P. Pelletier, and J. Martel-Pelletier, "Machine-learning-based patient-specific prediction models for knee osteoarthritis," *Nature Reviews Rheumatology*, vol. 15, no. 1, pp. 49–60, jan 2019.

[11] R. W. Telles, L. A. C. Machado, L. Costa-Silva, and S. M. Barreto, "Cohort profile update: The brazilian longitudinal study of adult health musculoskeletal (ELSA-Brasil MSK) cohort," *International Journal of Epidemiology*, vol. 51, no. 6, pp. e391–e400, dec 2022.

[12] R. W. Telles, L. Costa-Silva, L. A. C. Machado, R. C. P. D. Reis, and S. M. Barreto, "Fixed-flexion knee radiography using a new positioning device produced highly repeatable measurements of joint space width: ELSA-Brasil musculoskeletal study (ELSA-Brasil MSK)," *Revista Brasileira de Reumatologia (English Edition)*, vol. 57, no. 2, pp. 154–161, mar/apr 2017.

[13] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[14] Y. Zhang and J. M. Jordan, "Epidemiology of osteoarthritis," *Clinics in Geriatric Medicine*, vol. 26, no. 3, pp. 355–369, aug 2010, erratum in: Clin Geriatr Med. 2013 May;29(2):ix.

[15] V. Silverwood, M. Blagojevic-Bucknall, C. Jinks, J. L. Jordan, J. Protheroe, and K. P. Jordan, "Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis," *Osteoarthritis and Cartilage*, vol. 23, no. 4, pp. 507–515, apr 2015.

[16] N. Yoshimura, S. Muraki, H. Oka, S. Tanaka, H. Kawaguchi, K. Nakamura, and T. Akune, "Accumulation of metabolic risk factors such as overweight, hypertension, dyslipidaemia, and impaired glucose tolerance raises the risk of occurrence and progression of knee osteoarthritis: a 3-year follow-up of the ROAD study," *Osteoarthritis and Cartilage*, vol. 20, no. 11, pp. 1217–1226, nov 2012.

[17] M. I. Schmidt, B. B. Duncan, J. G. Mill, P. A. Lotufo, D. Chor, S. M. Barreto, E. M. Aquino, V. M. Passos, S. M. Matos, M. C. Molina, M. S. Carvalho, and I. M. Bensenor, "Cohort profile: Longitudinal study of adult health (ELSA-Brasil)," *International Journal of Epidemiology*, vol. 44, no. 1, pp. 68–75, feb 2015.

[18] K. J. Rothman, S. Greenland, and T. L. Lash, *Modern Epidemiology*, 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2013.

TABLE II: List of Selected Variables for the Model.

Category	Variable / Description
Osteoarthritis (OA) and Radiographic Assessment	
OA Classification	Participant-level, right knee, and left knee classifications.
Symptomatic OA Classification	Participant-level, right knee, and left knee classifications for symptomatic OA.
Kellgren-Lawrence (KL) Classification	For tibiofemoral (TF PA view) and patellofemoral (PF profile view) compartments of both the right and left knee.
Radiograph	General imaging variable.
Symptoms and Physical Function	
Pain	Frequent knee pain; Prevalent knee pain (last 7 days).
Function and Disability	WOMAC - physical function score (0-170); Knee disability; Repeated sit-to-stand test performance.
Body Composition and Muscle Strength	
Body Mass	Bone mineral content / Mineral mass (kg); Skeletal muscle mass (kg).
Muscle Condition	Sarcopenia; Categorized Handgrip Strength.
Clinical History and Mechanical Factors	
Trauma and Surgery	History of trauma/injury to the knees; History of surgery on the knees; Onset of pain due to trauma.
Demographic and Occupational Data	
Personal Data	Sex; Age (continuous, grouped); Year of Birth.
Occupation	Nature of occupation.
Anthropometric Data	
Body Mass Index (BMI)	Continuous (kg/m²) and categorized (4 and 6 categories).
Fat Distribution	Abdominal obesity; Waist-hip ratio.
Lifestyle and Habits	
Physical Activity	IPAQ score.
Smoking	Smoking habits.
Alcohol Consumption	General use, heavy drinking (male $\geq$ 210 g/week; female $\geq$ 140 g/week), and binge drinking.
Metabolic Health and Cardiovascular Risk	
Metabolic Syndrome Definition	Based on JIS criteria for different ethnicities (Central/South American, North American - ATPIII, and European descent - IDF).
Metabolic Syndrome Components	Systemic arterial hypertension; Hypertriglyceridemia (with and without considering medication); Low HDL-C levels (with and without considering medication); Impaired glucose tolerance / Diabetes mellitus (various definitions).
Framingham Risk Scores	Coronary Heart Disease (points for Cholesterol and LDL); Cardiovascular Disease (models for Cholesterol and BMI).