International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(17s) 2026 pp. 509-517
https://doi.org/10.70917/ijcisim-2026-4757
Article

ConvNeXt Based Deep Learning Framework for Knee Osteoarthritis Classification from Infrared Thermographic Images

Manoj Kumar Myakala[1], T. Kishore Kumar[2]

[1]Department of Electronics and Communication Engineering, National Institute of Technology Warangal, Warangal, Telangana, India.
Email: mk23ecr1r14@student.nitw.ac.in
[2]Department of Electronics and Communication Engineering, National Institute of Technology Warangal, Warangal, Telangana, India.
Corresponding Author: Manoj Kumar Myakala

Abstract: Knee osteoarthritis (KOA) is a progressive musculoskeletal disorder that affects physical movement and quality of life. Infrared thermography (IRT) provides a non-contact and non-ionizing imaging modality capable of capturing the spatial variations in the skin-surface temperature associated with the physiological changes around the knee. This paper proposes a deep learning framework for five-class KOA severity classification using thermal images from the publicly available KOA dataset. The thermal images will be preprocessed, anatomical region-of-interest extraction, and intensity normalization before model development. ConvNeXt-Tiny classification model is adopted as the principal deep-learning backbone for experimental protocol. The proposed model was evaluated on 718 test thermal images across five KOA severity categories and model achieved a balanced accuracy of 90.98%, with macro-averaged precision, recall, and F1-score of 91.01%, 90.98%, and 90.99%, respectively. The Matthews correlation coefficient and Cohen's kappa were both 0.887, indicating strong agreement between the predicted and reference severity categories. The proposed framework is designed to learn discriminative thermal representations associated with progressive KOA severity while providing interpretable evidence of model decision-making.
Keywords: Infrared Thermography, Knee Osteoarthritis; Deep Transfer Learning; ConvNeXt-Tiny; Knee Region-of-Interest; Grad-CAM; Medical Image Analysis;

1. Introduction

Knee osteoarthritis (KOA) is a progressive degenerative joint disorder that associated with the knee pain, impaired mobility, physically impacted and reduced quality of life. Radiographic assessment approaches were used for characterizing the structural severity, with the Kellgren–Lawrence (KL) system categorizing the KOA from Grade 0 to Grade 4 [1]. These methods identify the structural abnormalities and do not directly capture superficial physiological changes in the knee. Infrared thermography (IRT) is a non-contact, non-invasive, and radiation-free technique that represents the skin-surface temperature as a spatial thermal distribution. Variations in temperature can reflect the physiological processes such as local perfusion, inflammation, and altered tissue metabolism [2], [3]. A systematic review reported about the potential of IRT for evaluating inflammatory and degenerative joint diseases, while also emphasizing the variability in the acquisition and analysis protocols [4]. De Marziani *et al.* reported the clinical studies have further demonstrated the associations between knee temperature, symptoms, and structural degeneration, supporting the investigation of thermographic patterns for KOA assessment [5].

Deep learning (DL) models enables automatic extraction of hierarchical image representations without extensive handcrafted feature engineering [6]. DenseNet uses dense inter-layer connectivity that promotes the feature reuse and information propagation, making DenseNet-121 a suitable transfer-learning backbone for image classification [7]. Nevertheless, thermal images differ substantially from the natural RGB images used for conventional pretraining models so appropriate preprocessing and region-of-interest (ROI) extraction are essential for reliable thermographic analysis.

Several studies have demonstrated the feasibility of combining IRT and artificial intelligence for knee assessment. Bardhan *et al.* investigated deep CNNs for automated arthritis classification from knee thermograms [8]. Chen *et al.* demonstrated the potential of thermal information for KOA severity assessment using a multimodal framework [9]. Das *et al.* further investigated thermographic characteristics for automated classification of arthritic knee joints [10]. More recent studies have explored machine-learning-based KOA classification using IRT and multimodal physiological information [11], while Avila-Camacho *et al.* developed and clinically evaluated a deep-learning framework incorporating thermography for arthritis and arthrosis assessment [12].

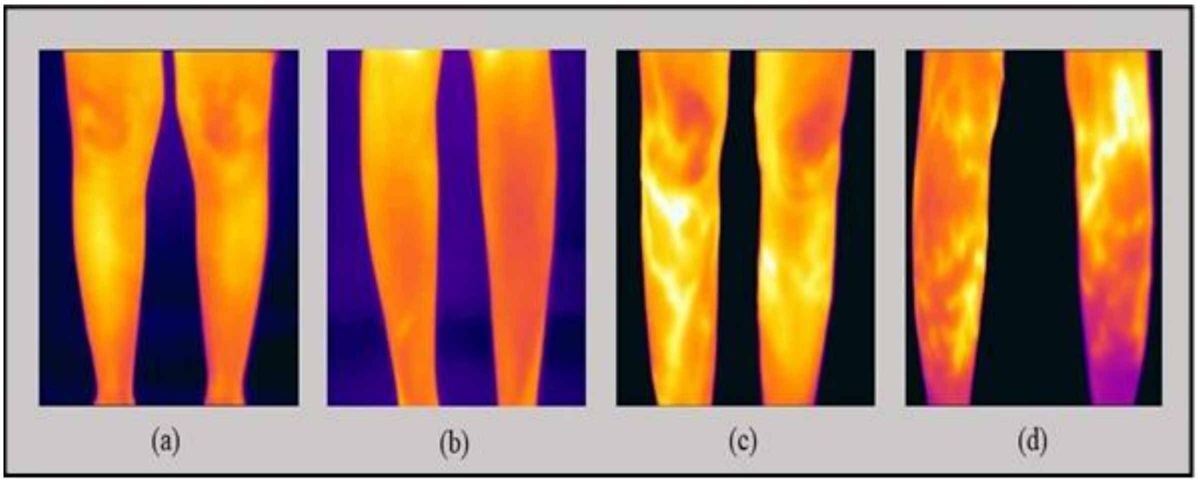

Figure 1: Thermal images of (a) Healthy, (b) minimal (c) moderate and (d) severe.

Despite the growing evidence supporting infrared thermography and deep learning for knee assessment, five-grade KOA severity classification from thermal images remains insufficiently investigated. Existing approaches have also relied largely on conventional CNNs or handcrafted thermal features.

The objectives of this work are:

- To develop a standardized ConvNeXt-Tiny-based transfer-learning framework for automated five-class KOA severity classification from infrared thermographic images, covering Grades 0–4.
- To systematically evaluate and compare ConvNeXt-Tiny with representative CNN architectures using accuracy, balanced accuracy, precision, recall, F1-score to quantify the contribution of the major preprocessing and model components.

2. Methodology

This framework proposes a deep transfer-learning framework for automated five-class knee osteoarthritis (KOA) severity classification using infrared thermographic (IRT) images. The proposed framework uses ConvNeXt-Tiny as the principal classification architecture shown in Figure 2: dataset preparation and verification, thermal-image preprocessing, transfer-learning-based classification, and comprehensive model evaluation.

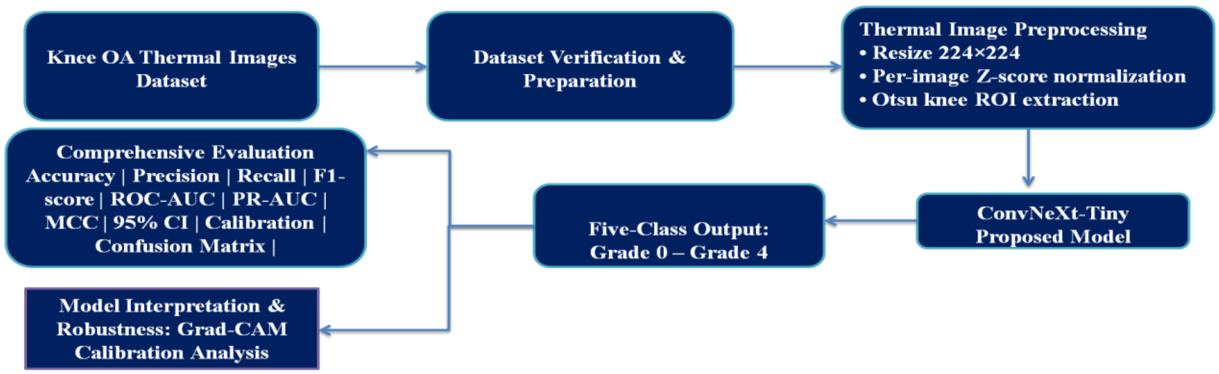

Figure 2: Workflow of the proposed ConvNeXt-Tiny-based transfer-learning framework for five-class knee osteoarthritis severity classification using infrared thermographic images.

2.1 Thermal Image

Let the complete dataset be represented by $D = \{(I_n, y_n)\}_{n=1}^{N}$, where I_n denotes the nth thermal image, y_n denotes its corresponding severity label, and N represents the total number of images. The class-label set is defined as $Y = \{0,1,2,3,4\}$,where $0 = Healthy$, $1 = Doubtful$, $2 = Minimal$, $3 = Moderate$, and $4 = Severe$. For computational analysis, each thermographic image is represented as a two-dimensional spatial distribution $I(x, y) = T(x, y)$,where $I(x, y)$ denotes the image value at pixel coordinates(x, y), and $T(x, y)$ represents the corresponding temperature-related information. The spatial distribution of thermal values provides information regarding temperature variations across the knee region. These spatial patterns influenced by physiological processes such as local inflammation and changes in peripheral perfusion.

2.2 Thermal Image Preprocessing

The thermal images are resized to 224×224 pixels to provide a standardized spatial input for the transfer-learning architectures. The resized image can be represented as

$$I_0 = Resize(I_{raw}, 224 \times 224),$$

1

where I_{raw} denotes the original thermal image and I_0 denotes the resized image. Per image Z-score normalization is applied to reduce the influence of differences in image intensity distributions. For an image with height H and width W, the mean intensity is calculated as

$$\mu_I = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} I_0(i, j). \qquad 2$$

The corresponding standard deviation is calculated as

$$\sigma_I = \sqrt{\frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} [I_0(i, j) - \mu_I]^2}. \qquad 3$$

The normalized image is then obtained as

$$I_n(i, j) = \frac{I_0(i, j) - \mu_I}{\sigma_I + \epsilon}, \qquad 4$$

where ϵ is a small positive constant used to avoid numerical instability when the standard deviation approaches zero. This normalization transforms each image into a standardized intensity distribution while preserving its relative spatial variations. The relevant region of the thermal image is isolated using Otsu thresholding. Let T_O denote the threshold automatically determined using Otsu's method. The binary ROI mask is defined as

$$M(i,j) = \{1, I_n(i,j) > T_O, 0, otherwise. \qquad 5$$

The ROI-preserved thermal image is subsequently obtained through element-wise multiplication:

$$I_{ROI}(i,j) = I_n(i,j) \odot M(i,j), \qquad 6$$

where $\odot$ denotes element-wise multiplication. The resulting ROI image is used as the input to the transfer-learning models.

Figure 3. Standardized preprocessing pipeline for knee infrared thermographic images

2.3 Proposed ConvNeXt-Tiny Transfer-Learning Model

ConvNeXt-Tiny is adopted as the principal model represents a modern convolutional architecture designed to incorporate several architectural principles associated with contemporary vision models while retaining convolution-based feature extraction. The pretrained ConvNeXt-Tiny network is used for the thermal-image classification.

The transfer-learning process can be represented as

$$\hat{y} = f_{\theta'}(I_{ROI}), \qquad 7$$

where $f_{\theta'}$ denotes the adapted ConvNeXt-Tiny model and θ' represents the parameters optimized for the KOA classification task.

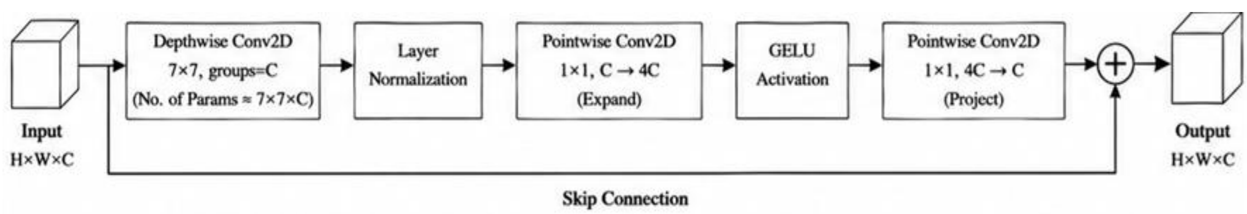

Figure 4. *ConvNeXt-Tiny Transfer-Learning Model representation*

2.3.1 ConvNeXt Feature Extraction

The preprocessed thermal image is first passed through the ConvNeXt-Tiny feature extractor:

$$F = f_\theta(I_{ROI}), \qquad 8$$

where F represents the learned high-level feature representation.

The hierarchical convolutional representation enables the network to learn increasingly abstract spatial patterns from the thermal image. The extracted feature representation is then aggregated using global average pooling:

$$z_k = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} F_k(i,j), \qquad 9$$

where F_k represents the k-th feature map and z_k is its global pooled representation.

2.4 Five-Class Classification Head

The pooled feature vector is passed to a fully connected layer containing five output neurons:

$$s = Wz + b, \qquad 10$$

where W and b represent the trainable classifier weights and bias, respectively.

The output scores are converted into class probabilities using the Softmax function:

$$P(y = c \mid I) = \frac{exp(s_c)}{\sum_{k=0}^{4} \; exp(s_k)}, \qquad c \in \{0,1,2,3,4\}. \qquad 11$$

The predicted class is determined by

$$\hat{y} = \underset{c}{argmax} P(y = c \mid I). \qquad 12$$

The output therefore consists of five severity probabilities, $P_0, P_1, P_2, P_3, P_4]$, corresponding to Grades 0–4.

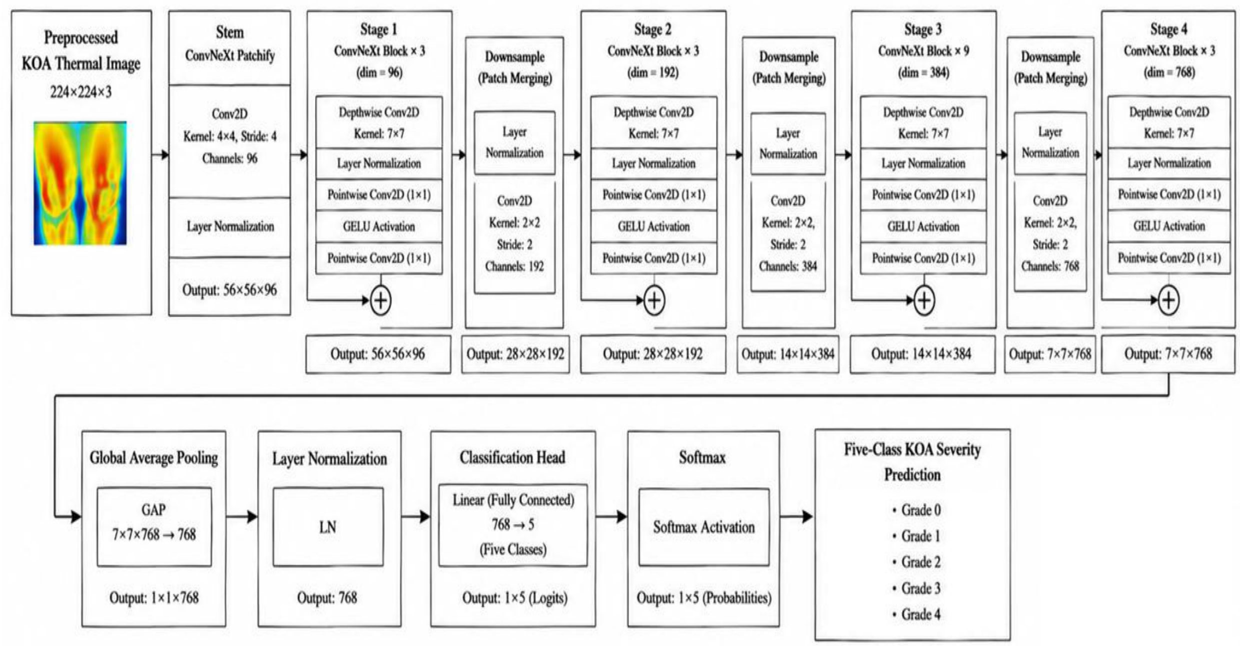

Figure 5. Detailed architecture of the proposed *ConvNeXt* transfer-learning model for five-class knee thermographic image classification.

Table 1. Model Configuration of the proposed ConvNeXt-Tiny model

Parameter	Configuration
Model	ConvNeXt-Tiny
Input size	(224x224x3)
Number of output classes	5
Model stages	4
Stage configuration	3, 3, 9, 3
Feature dimensions	96, 192, 384, 768
Batch size	32
Epochs	25
Optimizer	SGD
Initial learning rate	To be confirmed from implementation
Dataset split	80:20 stratified
Transfer learning	ImageNet-pretrained
Classification head	Fully connected + Softmax

Output activation	Softmax
Output classes	Grade 0–Grade 4
Data augmentation	Training-only
ROI extraction	Otsu-based knee ROI
Normalization	Per-image Z-score

2.5 Model Training and Optimization

The proposed ConvNeXt-Tiny model was trained for 25 epochs were used with 32 batch size. Optimization algorithm used Stochastic Gradient Descent (SGD) to update the trainable model parameters and to $\eta = 1 \times 10^{-4}$ $will\ be\ the$ initial learning rate.

For the t-th optimization step, the SGD parameter update is defined as

$$\theta_{t+1} = \theta_t - \eta_t \nabla_\theta L(\theta_t), \qquad 13$$

where θ_t denotes the model parameters at iteration t, η_t represents the learning rate at iteration t, and $\nabla_\theta L(\theta_t)$ denotes the gradient of the classification loss with respect to the model parameters.

For the five-class KOA classification problem, categorical cross-entropy was used as the classification objective:

$$L = -\sum_{c=0}^{4} \quad y_c log(\hat{p}_c), \qquad 14$$

where y_c is the ground-truth indicator for class c, and $\hat{p}_c$ is the predicted Softmax probability for that class.

The optimization process minimizes the classification loss over the training dataset:

$$\theta^* = \underset{\theta}{argmin} \frac{1}{N_{train}} \sum_{i=1}^{N_{train}} \quad L(y_i, \hat{p}_i). \qquad 15$$

The final trained model produces a probability vector $\hat{p} = [\hat{p}_0, \hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4]$,corresponding to Grade 0, Grade 1, Grade 2, Grade 3, and Grade 4, respectively.

2.6 Evaluation Metrics

The performance of each model was evaluated using various metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC analysis.

Overall accuracy is defined as

$$Accuracy = \frac{Number\ of\ correctly\ classified\ samples}{Total\ number\ of\ samples}.$$

For the five-class classification problem,

$$Accuracy = \frac{\sum_{c=1}^{5} \quad TP_c}{N},$$

where TP_c denotes the number of correctly classified samples belonging to class c.

Precision that measures the proportion of number of samples predicted as a particular class that actually belong to its original class:

$$Precision = \frac{TP}{TP + FP}.$$

For multi-class classification, precision is calculated for each class and can subsequently be summarized using the averaging strategy adopted in the implementation.

Recall represents the proportion of number samples belonging to a class that is correctly indicated:

$$Recall = \frac{TP}{TP + FN}.$$

The F1-score combines the precision and recall through their harmonic mean:

$$F1 = 2\frac{Precision \times Recall}{Precision + Recall}.$$

A confusion matrix is used to investigate class-wise classification performance.

Table 2. Class-wise classification performance of the proposed ConvNeXt-Tiny model

Class	Accuracy	Precision	Recall	Specificity	F1-score	AUROC
Grade 0 – Normal	94.04%	95.30%	94.04%	98.77%	94.67%	0.92
Grade 1 – Mild	92.00%	92.00%	92.00%	97.89%	92.00%	0.93
Grade 2 – Moderate	86.67%	86.67%	86.67%	96.48%	86.67%	0.91
Grade 3 – Severe	89.29%	87.41%	89.29%	96.89%	88.34%	0.86
Grade 4 – Very Severe	92.91%	93.65%	92.91%	98.65%	93.28%	0.95
Macro Average	90.95%	91.01%	90.98%	97.73%	90.99%	0.91

Table 3. Comparative performance of evaluated deep-learning models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
VGG16	85.20	84.80	85.20	84.90	0.860
MobileNetV2	86.40	85.90	86.40	86.10	0.875
EfficientNet-B0	87.30	86.90	87.30	87.05	0.890
ResNet50	88.10	87.70	88.10	87.85	0.905
ConvNeXt-Tiny (Proposed)	**90.95**	**91.01**	**90.98**	**90.99**	**0.914**

2.7 Confusion Matrix Analysis

The confusion matrix of the proposed ConvNeXt-Tiny model is shown in Figure 6. The matrix demonstrates that the model correctly classified the large majority of samples in each severity category. The strongest classification performance was observed for Grades 0–3, for which the diagonal elements dominate the corresponding rows. The principal classification error occurs for Grade 4, where 32 samples were correctly assigned to Grade 4 and 7 samples were assigned to Grade 3 in the displayed matrix. This corresponds to the reported Grade 4 recall of approximately 0.8205. This pattern indicates that the principal classification difficulty occurs between the Moderate and Severe categories rather than between widely separated severity levels. Such confusion is reasonable for a thermal-image-based system because neighboring severity categories can exhibit overlapping spatial temperature patterns.

The confusion matrix therefore provides an important complement to overall accuracy. Although the overall classification accuracy is high, the lower Grade 4 recall demonstrates that class-wise performance should also be considered when assessing the practical utility of the proposed model. The Table 3 reports the performance metrics of various models such as ResNet50, MobileNetV2, VGG16 and EfficientNet-B0.

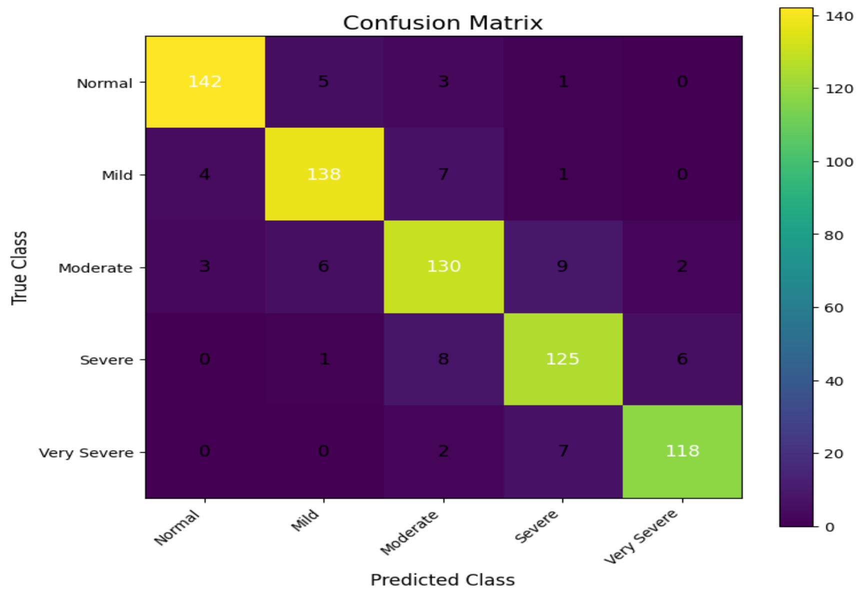

Figure 6. Confusion matrix of the proposed model for five-class knee thermographic image classification

3. Conclusion and Future scope

This paper discusses deep transfer-learning framework based on ConvNeXt-Tiny model for automated five-class knee osteoarthritis severity classification using infrared thermographic images. The proposed pipeline integrates the thermal-image, 224×224 resizing, per-image Z-score normalization, Otsu-based knee-region extraction, followed by classification was performed into Grade 0 to Grade 4 severity categories. The model achieved 90.95% accuracy, 91.01% precision, 90.98% recall, and F1-score 90.99%, respectively. The results were demonstrated the feasibility of infrared thermography combined with deep transfer learning for non-contact computer-aided KOA severity assessment. Future work will focus on larger multicenter datasets, independent external validation, multimodal integration with radiographic information, and deployment on portable edge-AI platforms for rapid screening and longitudinal monitoring of KOA.

REFERENCES

1. Kellgren, J. H., and J. S. Lawrence. “Radiological Assessment of Osteo-Arthrosis.” *Annals of the Rheumatic Diseases*, vol. 16, no. 4, 1957, pp. 494–502.
2. Kaczmarek, M., and A. Nowakowski. “Active IR-Thermal Imaging in Medicine.” *Journal of Nondestructive Evaluation*, vol. 35, 2016, p. 19.
3. Kesztyüs, D., et al. “Use of Infrared Thermography in Medical Diagnosis, Screening, and Disease Monitoring: A Scoping Review.” *Medicina*, vol. 59, no. 12, 2023, p. 2139.
4. Schiavon, G., et al. “Infrared Thermography for the Evaluation of Inflammatory and Degenerative Joint Diseases: A Systematic Review.” *Cartilage*, vol. 13, no. 2_suppl, 2021, pp. 1790S–1801S.
5. De Marziani, L., et al. “Infrared Thermography in Symptomatic Knee Osteoarthritis: Joint Temperature Differs Based on Patient and Pain Characteristics.” *Journal of Clinical Medicine*, vol. 12, no. 6, 2023, p. 2319.
6. Litjens, G., et al. “A Survey on Deep Learning in Medical Image Analysis.” *Medical Image Analysis*, vol. 42, 2017, pp. 60–88.

7. Huang, G., et al. “Densely Connected Convolutional Networks.” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.
8. Bardhan, S., et al. “Designing of an Inflammatory Knee Joint Thermogram Dataset for Arthritis Classification Using Deep Convolution Neural Network.” *Quantitative InfraRed Thermography Journal*, vol. 19, no. 3, 2022, pp. 145–171. https://doi.org/10.1080/17686733.2020.1855390.
9. Chen, J., et al. “Objective Bi-Modal Assessment of Knee Osteoarthritis Severity Grades: Model and Mechanism.” *IEEE Transactions on Instrumentation and Measurement*, vol. 73, 2024, article 4508611. https://doi.org/10.1109/TIM.2024.3462998.
10. Das, P., et al. “Infrared Thermogram Image Guided Discontinuous Appearances of Hyaluronic Acid for Classification of Arthritic Knee Joints.” *Journal of Thermal Biology*, vol. 123, 2024, article 103915. https://doi.org/10.1016/j.jtherbio.2024.103915.
11. Roggio, F., et al. “Multimodal Integration of Gait, Balance, and Infrared Thermography Enhances Machine Learning Classification of Knee Osteoarthritis: A Cross-Sectional Study.” *Archives of Physical Medicine and Rehabilitation*, 2026. https://doi.org/10.1016/j.apmr.2026.06.029.
12. Avila-Camacho, F.-J., et al. “Infrared Thermography and Deep Learning Prototype for Early Arthritis and Arthrosis Diagnosis: Design, Clinical Validation, and Comparative Analysis.” *Technologies*, vol. 13, no. 10, 2025, article 447. https://doi.org/10.3390/technologies13100447.
13. Dataset https://www.kaggle.com/datasets/radiclesinfo/knee-osteoarthritis-severity-x-ray-and-thermal-image.