technologies

MDPI

Article

Infrared Thermography and Deep Learning Prototype for Early Arthritis and Arthrosis Diagnosis: Design, Clinical Validation, and Comparative Analysis

Francisco-Jacob Avila-Camacho [1,*], Leonardo-Miguel Moreno-Villalba [2], José-Luis Cortes-Altamirano [3,4], Alfonso Alfaro-Rodríguez [3], Hugo-Nathanael Lara-Figueroa [5], María-Elizabeth Herrera-López [4] and Pablo Romero-Morelos [4]

1 División de Ingeniería en Sistemas Computacionales, TecNM/Tecnológico de Estudios Superiores de Ecatepec, Ecatepec de Morelos 55210, Estado de México, Mexico
2 División de Ingeniería Informática, TecNM/Tecnológico de Estudios Superiores de Ecatepec, Ecatepec de Morelos 55210, Estado de México, Mexico; lmoreno@tese.edu.mx
3 Instituto Nacional de Rehabilitación Luis Guillermo Ibarra, Ciudad de México 14389, Mexico; drjlcortesaltamirano@gmail.com (J.-L.C.-A.); alfa1360@yahoo.com.mx (A.A.-R.)
4 Licenciatura en Quiropráctica, Universidad Estatal del Valle de Ecatepec, Ecatepec de Morelos 55210, Estado de México, Mexico; eliza_herrera83@yahoo.com.mx (M.-E.H.-L.); pablo.r.morelos@gmail.com (P.R.-M.)
5 División de Ingeniería en Gestión Empresarial, TecNM/Tecnológico de Estudios Superiores de Ecatepec, Ecatepec de Morelos 55210, Estado de México, Mexico; hugolara@tese.edu.mx
* Correspondence: fjacobavila@tese.edu.mx

Abstract

Arthritis and arthrosis are prevalent joint diseases that cause pain and disability, and their early diagnosis is crucial for preventing irreversible damage. Conventional diagnostic methods such as X-ray, ultrasound, and MRI have limitations in early detection, prompting interest in alternative techniques. This work presents the design and clinical evaluation of a prototype device for non-invasive early diagnosis of arthritis (inflammatory joint disease) and arthrosis (osteoarthritis) using infrared thermography and deep neural networks. The portable prototype integrates a Raspberry Pi 4 microcomputer, an infrared thermal camera, and a touchscreen interface, all housed in a 3D-printed PLA enclosure. A custom Flask-based application enables two operational modes: (1) thermal image acquisition for training data collection, and (2) automated diagnosis using a pre-trained ResNet50 deep learning model. A clinical study was conducted at a university clinic in a temperature-controlled environment with 100 subjects (70% with arthritic conditions and 30% healthy). Thermal images of both hands (four images per hand) were captured for each participant, and all patients provided informed consent. The ResNet50 model was trained to classify three classes (healthy, arthritis, and arthrosis) from these images. Results show that the system can effectively distinguish healthy individuals from those with joint pathologies, achieving an overall test accuracy of approximately 64%. The model identified healthy hands with high confidence (100% sensitivity for the healthy class), but it struggled to differentiate between arthritis and arthrosis, often misclassifying one as the other. The prototype's multiclass ROC (Receiver Operating Characteristic) analysis further showed excellent discrimination between healthy vs. diseased groups (AUC, Area Under the Curve ~1.00), but lower performance between arthrosis and arthritis classes (AUC ~0.60–0.68). Despite these challenges, the device demonstrates the feasibility of AI-assisted thermographic screening: it is completely non-invasive, radiation-free, and low-cost, providing results in real-time. In the discussion, we compare this thermography-based approach with conventional diagnostic modalities and highlight its advantages, such as early detection of physiological changes, portability, and patient comfort. While not intended to replace established methods, this

check for updates

Academic Editor: Lipo Wang

Received: 7 August 2025
Revised: 16 September 2025
Accepted: 16 September 2025
Published: 2 October 2025

Citation: Avila-Camacho, F.-J.; Moreno-Villalba, L.-M.; Cortes-Altamirano, J.-L.; Alfaro-Rodríguez, A.; Lara-Figueroa, H.-N.; Herrera-López, M.-E.; Romero-Morelos, P. Infrared Thermography and Deep Learning Prototype for Early Arthritis and Arthrosis Diagnosis: Design, Clinical Validation, and Comparative Analysis. *Technologies* **2025**, *13*, 447. https://doi.org/10.3390/technologies13100447

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

technology can serve as an early warning and triage tool in clinical settings. In conclusion, the proposed prototype represents an innovative application of infrared thermography and deep learning for joint disease screening. With further improvements in classification accuracy and broader validation, such systems could significantly augment current clinical practice by enabling rapid and non-invasive early diagnosis of arthritis and arthrosis.

Keywords: infrared thermography; arthritis diagnosis; deep learning

1. Introduction

Arthritis (inflammatory joint disease, including rheumatoid arthritis) and arthrosis (osteoarthritis, a degenerative joint disease) are common musculoskeletal conditions that progressively impair joint function. Osteoarthritis (OA) is the most prevalent musculoskeletal disease, affecting roughly 20–30% of the adult population in developed countries [1], and is a leading cause of pain and disability worldwide. Rheumatoid arthritis (RA), while less common (prevalence ~0.5–1% [1]), is a serious autoimmune condition that causes chronic joint inflammation and can lead to severe disability if untreated. Both conditions significantly reduce quality of life and work capacity for patients. Early diagnosis and intervention are critical. In RA, prompt treatment in the initial stages can prevent irreversible joint damage and deformities. In OA, early lifestyle or therapeutic interventions may slow disease progression. However, diagnosing these conditions at an incipient stage remains challenging with traditional methods.

Conventional diagnostic techniques for arthritis and arthrosis include clinical evaluation and imaging. X-ray radiography is commonly used, especially for osteoarthritis, to detect joint space narrowing, osteophytes, and bone erosion. However, these structural changes typically manifest only in moderate to advanced stages of the disease, making X-rays insensitive for early diagnosis when no irreversible damage is yet present [2]. In the case of RA, radiographic changes may lag behind clinical symptoms by months, delaying timely intervention. Magnetic resonance imaging (MRI) offers higher sensitivity for detecting early inflammatory signs, such as synovitis and bone marrow edema, allowing detection before structural damage occurs. Nevertheless, MRI is limited by its high cost, limited availability, long acquisition times, and patient discomfort, making it impractical for routine screening or frequent monitoring [3]. Musculoskeletal ultrasound is more accessible and can dynamically visualize synovial thickening and effusion; however, its accuracy is highly dependent on the operator and may vary between practitioners. Additionally, while serological markers (e.g., rheumatoid factor, anti-CCP) support RA diagnosis, they lack localization and are not always positive in early disease. These limitations highlight the need for non-invasive, rapid, and accessible tools capable of detecting physiological changes—such as increased metabolic activity and perfusion—before irreversible structural damage occurs [4].

Infrared thermography has emerged as a promising adjunct diagnostic technique for this purpose [5]. Thermography is a non-contact imaging modality that captures the surface temperature distribution of the body, reflecting underlying blood flow and metabolic activity. In the context of joint diseases, inflammation is a key pathological feature; inflamed joints exhibit increased blood perfusion and heat emission [6]. A rise in skin temperature over a joint can thus serve as an indicator of active arthritis [1,7]. Thermographic evaluation of patients with rheumatic diseases has shown that thermal patterns correlate with disease presence, severity, and even treatment responses [7]. A recent systematic review concluded that infrared thermography is a simple, accurate, noninvasive, and radiation-free method

that can supplement conventional tools for screening and monitoring of both rheumatoid arthritis and osteoarthritis [1,8].

Unlike X-rays or MRI, thermography does not visualize anatomical structures but rather physiological abnormalities, detecting inflammation even in the absence of visible structural damage [9]. Moreover, as an imaging technique, it is completely safe; it does not involve ionizing radiation or any contact, posing no risk or discomfort to the patient [10]. This makes it suitable for frequent use and for vulnerable populations. Thermography is also cost-effective; modern infrared cameras are relatively affordable compared to hospital imaging equipment [11]. However, historically, medical thermography faced skepticism due to variability and the difficulty of interpreting thermal images [12]. Recent advances in sensor technology and computational analysis (especially artificial intelligence) are overcoming these challenges [13]. High-resolution thermal sensors and standardized imaging protocols have improved data quality, while machine learning algorithms can automatically analyze complex thermal patterns, increasing diagnostic accuracy [14]. In fields such as oncology, AI-assisted thermography is already demonstrating impressive results. For instance, Yousefi et al. [11] developed an embedded deep learning system for real-time thermographic analysis in breast cancer screening, highlighting a paradigm shift toward point-of-care, low-cost, and portable diagnostic devices. Similarly, in rheumatology, AI-enhanced thermography has shown promise in detecting inflammatory conditions with high sensitivity [4]. These developments suggest that thermography, when combined with edge-compatible AI models, can transition from a research tool to a practical clinical solution—particularly for early screening in resource-limited settings [15,16]. In this context, our work aligns with this emerging trend by proposing a fully integrated, low-cost prototype that performs on-device inference, enabling real-time, non-invasive joint disease screening without reliance on cloud computing or specialized infrastructure.

Several recent studies have explored the use of infrared thermography in diagnosing musculoskeletal disorders, particularly arthritis. Authors of [1] conducted a systematic review and concluded that thermographic analysis of joint areas correlates with the severity of both rheumatoid arthritis and osteoarthritis, especially when combined with quantitative feature extraction. Ref. [17] emphasized the potential of AI-enhanced thermography in medical screening, highlighting cases in oncology where machine learning models reached diagnostic accuracy near that of traditional modalities such as mammography [6]. In rheumatology, Ref. [8] presented results using handheld thermal cameras combined with machine learning models to classify arthritis and healthy patients, achieving over 80% accuracy in binary classification. These findings suggest that thermography, when integrated with AI, holds promise not only for detecting joint inflammation but also for facilitating fast and accessible screening solutions. However, few studies have implemented standalone, portable systems with integrated diagnostic capabilities, particularly ones aimed at multi-class classification (healthy, arthritis, arthrosis). Our work seeks to address this gap by developing and evaluating a self-contained prototype designed for early joint disease detection in clinical settings.

In recent years, significant advancements have been made in applying deep learning mechanisms such as attention modules and multimodal (vision–language) architectures to medical image analysis. Li et al. [18] presented a comprehensive review of attention mechanisms in medical imaging, highlighting their capacity to enhance model performance across classification, segmentation, and detection tasks by enabling networks to focus on clinically relevant regions, thereby improving both interpretability and accuracy [18]. Advances in segmentation tasks have been demonstrated via architectures such as MEDUSA, which incorporates multi-scale self-attention within a unified encoder–decoder framework and achieves state-of-the-art results on challenging benchmarks [19].

Simultaneously, the rise of vision–language models (VLMs) is reshaping multimodal medical AI. [18] reviewed the evolution of VLMs in medical imaging, charting their trajectory from simple fusion-based methods to more general large-scale models capable of integrating visual inputs with textual and clinical context [20,21]. Additionally, recent work such as BiPVL-Seg illustrates how bidirectional progressive fusion between vision and language modalities, with global–local alignment, can significantly enhance performance in medical image segmentation tasks [22].

Compared with these advanced AI methodologies, prior thermography-based diagnostic systems have not leveraged attention mechanisms or multimodal features to the same extent. Our prototype bridges this gap by integrating thermal imaging with deep learning in a streamlined, standalone, Flask-based device, providing not only early detection capabilities but also a user-friendly interface and portability—features rarely combined in existing solutions.

In this study, we apply infrared thermography and deep learning to the domain of joint disease diagnosis. We describe the design and construction of a prototype device intended as a point-of-care tool for early detection of arthritis and arthrosis. The prototype integrates a Raspberry Pi-based thermal imaging system with a custom deep neural network model to classify thermal images of patients' hands. We conducted clinical validation with real patients to evaluate the system's performance in distinguishing healthy individuals from those with arthritic conditions. The results are analyzed in terms of accuracy, confusion matrix, and ROC curves to assess how well the device can identify not only the presence of joint disease but also differentiate between inflammatory arthritis and osteoarthritis. Furthermore, we include a comparative discussion against existing diagnostic technologies, highlighting the benefits (and limitations) of AI-assisted thermography as an early diagnostic modality. The aim of this work is to demonstrate the feasibility of a low-cost, non-invasive screening tool for arthritis and arthrosis, and to position it in the context of current clinical practice and future technological trends.

2. Materials and Methods

2.1. Prototype Design

The diagnostic prototype was developed as a compact, self-contained unit centered around a Raspberry Pi 4 microcomputer (Raspberry Pi Foundation, Cambridge, UK). All components are enclosed in a custom housing 3D-printed with polylactic acid (PLA) thermoplastic (Formfutura, Valkenswaard, The Ntherlands), designed to be lightweight and portable. The device includes an infrared thermal camera module mounted at the front of the enclosure to capture thermal images of a patient's hands. The specific thermal camera used (FLIR Lepton V 2.0) (Teledyne FLIR Systems, Wilsonvile, OR, USA) provides a two-dimensional heat map (thermogram) of the scene, it is sensitive to long-wave infrared radiation, translating surface temperature variations into a false-color or grayscale image. The resolution of the thermal sensor is sufficient to discern temperature differences across small regions of the hand (on the order of a few millimeters). The camera was interfaced with the Raspberry Pi via a digital SPI/I^2C connection. For user interaction, a touch-sensitive display is embedded on the front panel of the device. This touch screen, 7 inches diagonal, serves both as a live viewfinder for image capture and as a graphical user interface for operating the system, Figure 1. The entire setup is powered by the Raspberry Pi's 5 V supply (with an external AC adapter and a battery pack) (Anker PowerCore 10000, Changsha, China), making the unit deployable at bedside or in various clinic rooms. To ensure reproducibility and provide a clear understanding of the system's capabilities, Table 1 summarizes the key technical specifications of the primary hardware components used in the prototype.



(**a**) (**b**)

Figure 1. Prototype designed and constructed as a portable diagnostic device for arthritis and arthrosis. (**a**) Front view; (**b**) Rear view, showing the location of the thermal camera FLIR Lepton.

Table 1. Technical specifications of the prototype's main components.

Component	Specification
FLIR Lepton v2.0	Resolution: 80 × 60 pixels; Thermal sensitivity: <50 mK (NETD); Spectral range: 7.5–13.5 µm; Frame rate: 9 Hz; Interface: SPI/I; Field of view (FoV): 57° × 45° (diagonal)
Raspberry Pi 4	CPU: Quad-core ARM Cortex-A72 (1.5 GHz) (Broadcom Limited, San Jose, CA, USA); RAM: 4 GB LPDDR4 (Micron Technology Inc., Boise, ID, USA); GPU: VideoCore VI; Storage: microSD card (32 GB) (Broadcom Limited, San Jose, CA, USA); OS: Raspberry Pi OS (64-bit) (Version 4 2023-10-11, Raspberry PI Foundation, Cambridge, United Kingdom); Power: 5 V/3 A
Touchscreen Display	Size: 7 inches; Resolution: 1024 × 600 pixels; Interface: HDMI + USB (for touch)
Power Supply	External adapter: 5 V/3 A; Portable battery (Anker PowerCore 10000, Changsha, China): 10,000 mAh, 5 V USB output

By using widely available components (the Raspberry Pi and a low-cost IR camera), the hardware cost of the prototype is kept relatively low, aligning with the goal of affordability.

2.2. Software and Application

A custom software application was developed using the Flask microframework (Python) (version 2.3.2, Pallets Projects, Portland, OR, USA) to run on the Raspberry Pi. The application provides a simple, touch-friendly interface with two main operational modes:

- Image Capture Mode (Training Data Acquisition): In this mode, the device can capture and save thermal images for building a dataset. Clinicians or researchers can use it to collect images from patients with known conditions. The interface allows the operator to capture a thermal image of the patient's hands and enter relevant metadata (such as the patient's ID and diagnosis class:—healthy, arthritis, or arthrosis—as well as gender and age). Four images per hand can be captured sequentially from different viewpoints or positions (e.g., dorsal and palmar views, or varying hand orientations) to comprehensively record the thermal profile. These images are stored in the device's memory, along with their labels, for later model training. The capture mode ensures

consistency by providing real-time thermal visualization and guiding the user on proper positioning (for example, an outline on-screen might indicate where to place the hand at a fixed distance from the camera to maintain uniform image framing).

- Diagnosis Mode (Automated Inference): In this mode, the system utilizes a pre-trained deep learning model to classify a new thermal image of a patient's hands and provide an instantaneous diagnostic suggestion. When diagnosis mode is activated, the user is prompted via the touchscreen to position the patient's hands in view of the camera (the device is placed at a fixed distance and angle, which was determined during calibration). Upon capturing the thermal image, the software preprocesses it and runs the image through the trained neural network. The predicted class (healthy, arthritis, or arthrosis) is then displayed on the screen, along with a confidence score or probability for transparency. The interface is designed for clarity in a clinical setting—for example, it might show a message like "Prediction: Arthritis (82% confidence)" on a clean background. An option to save the diagnostic image and result for record-keeping is also provided.

2.3. Deep Learning Model (ResNet50) and Training

The diagnostic algorithm is based on a convolutional neural network (CNN) using the ResNet50 architecture. While more recent models (e.g., EfficientNet, Vision Transformers) offer higher accuracy in some domains, ResNet50 was selected due to its favorable balance between performance and computational efficiency, which is critical for deployment on resource-constrained edge devices such as the Raspberry Pi 4 (4 GB RAM, quad-core ARM Cortex-A72). Given the goal of real-time, on-device inference without reliance on cloud connectivity, we prioritized a model with a manageable memory footprint and inference latency. ResNet50, pre-trained on ImageNet, provides strong feature extraction capabilities and is well-suited for transfer learning with limited medical imaging data. Its residual connections mitigate the vanishing gradient problem, enabling stable training even when fine-tuning. In contrast, deeper or more complex architectures (e.g., EfficientNet-B4 or ViT) require significantly more computational resources and longer inference times, making them less suitable for low-power, portable diagnostic systems. Therefore, ResNet50 represents a pragmatic and effective choice for this embedded AI application, where reliability, speed, and hardware compatibility are prioritized alongside accuracy.

The dataset for training the model was composed of thermal images collected from the clinical study (described below). Prior to input into the network, each thermal image was preprocessed: we normalized the temperature values and resized or padded the images to the input size expected by ResNet50 (224×224 pixels). In order to match the three-channel input of the pre-trained model, the single-channel thermal data were duplicated or mapped to a pseudo-RGB format (alternatively, a colormap could be applied to the thermal image to produce three channels). Data augmentation was applied on the fly to enrich the training set and improve model generalization, while strictly preserving the thermal integrity of the images. The techniques used included horizontal flipping (to account for left/right hand symmetry), small rotations ($\pm 10^{\circ}$), translations ($\pm 15\%$ of image width), and scaling (0.9–1.1×) to simulate variations in hand positioning, orientation, and distance from the camera. Crucially, no augmentation altered the pixel intensity (temperature) values beyond the initial normalization step. Temperature normalization was performed per image using min–max scaling:

$$T_{norm} = \frac{T - T_{min}}{T_{max} - T_{min}} \quad (1)$$

where T is the raw pixel temperature in °C. This normalization ensures that thermal contrasts—key indicators of inflammation—are preserved relative to the individual image,

while making the model robust to absolute temperature shifts caused by ambient conditions or camera calibration drift. By avoiding any artificial modification of thermal values (e.g., adding noise or contrast adjustments), we ensure that the augmented data reflects realistic variations in spatial presentation without introducing thermal artifacts. This approach enhances the model's robustness to minor environmental fluctuations while maintaining physiological fidelity in the thermal signal.

Min–max image normalization preserves relative thermal contrasts, which is critical for detecting focal inflammation, and avoids bias from absolute temperature differences due to ambient conditions or sensor calibration [8]. For image acquisition, a protocol was strictly designed and followed, avoiding environmental conditions that could affect patients' temperatures. This protocol is explained in Section 2.4.

The pre-trained layers of ResNet50 were initially frozen and later fine-tuned using a lower learning rate of 1×10^{-5}, while the newly added layers were trained with a learning rate of 1×10^{-4}. We applied a ReduceLROnPlateau scheduler to decrease the learning rate when validation loss plateaued, with a patience of 5 epochs and a reduction factor of 0.1.

The model was trained using the cross-entropy loss function to measure classification error. We employed the Adam optimizer (with an initial learning rate on the order of 1×10^{-5} for the new layers, and a lower rate for the pre-trained layers—1×10^{-4}) to update the network weights. Training was conducted for a sufficient number of epochs to ensure convergence; we empirically chose up to 200–250 epochs, monitoring performance on a validation set after each epoch. A small portion of the training data (10–15%) was set aside as a validation set to tune hyperparameters and perform early stopping. The training and validation loss and accuracy were tracked over epochs to diagnose overfitting or underfitting. As shown later in the Results section, the network's training progressed with steadily decreasing loss and increasing accuracy, eventually reaching a high accuracy on the validation set, suggesting the model learned to classify the training images effectively. The complete workflow from image capture to diagnosis is illustrated in Figure 2.

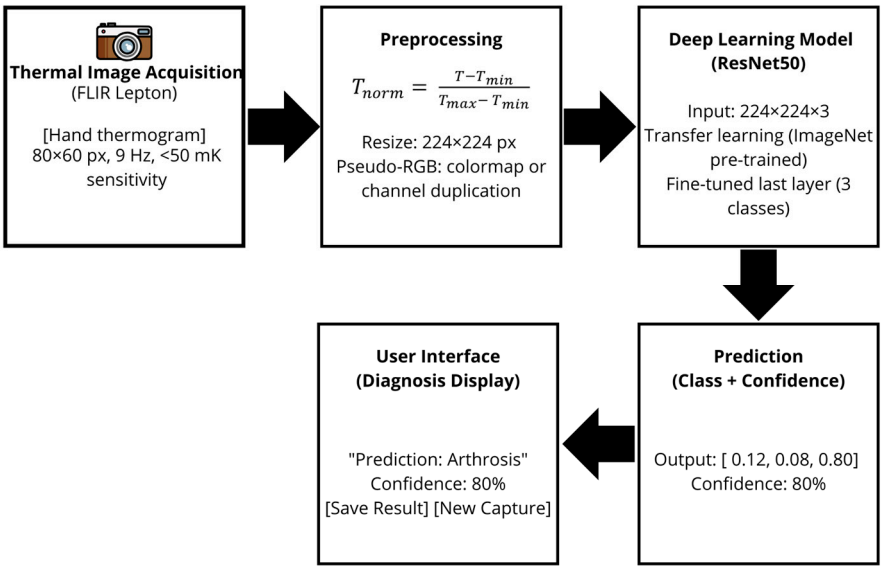


Figure 2. System workflow from thermal image acquisition to AI-based diagnosis.

The deep learning model was implemented using TensorFlow 2.18 with the Keras 3.7 API. Training was performed on a workstation equipped with an NVIDIA RTX 3060 GPU (NVIDIA Corporation, Santa Clara, CA, USA), leveraging mixed-precision training (float16/float32) to accelerate computation and reduce memory usage. For data preprocessing, augmentation, and model evaluation, we used standard Python libraries including

NumPy (2.0.2), OpenCV (4.12), and scikit-learn (v1.3). Specifically, scikit-learn was employed to compute the confusion matrix, classification report (precision, recall, F1 score), and multi-class ROC curves. All code was integrated into the Flask-based application running on the Raspberry Pi 4 for on-device inference.

2.4. Clinical Data Collection

A clinical validation study was carried out at the Universidad Estatal del Valle de Ecatepec (UNEVE) University Clinic with approval from the institution's ethics committee for the data collection protocol. A total of 100 volunteers were recruited for thermal imaging of their hands. These included patients diagnosed with arthritic conditions as well as healthy individuals serving as controls. Specifically, 70% of the participants had a form of clinically confirmed joint disease (either inflammatory arthritis or osteoarthritis), and the remaining 30% were healthy subjects without any known joint pathology. Within the diseased group, patients were categorized by a specialist based on clinical and radiological criteria: those with rheumatoid or other inflammatory arthritis vs. those with osteoarthritis ("artrosis" in Spanish). For the purposes of model training, we treated these diagnoses as two distinct target classes ("arthritis" and "arthrosis"). All participants provided written informed consent after being informed about the study's purpose and the non-invasive nature of thermographic imaging.

A total of 270 infrared images were collected from 100 participants at the UNEVE University Clinic using a controlled acquisition protocol. Each participant contributed four images per hand, resulting in 8 images per subject.

Thermal image acquisition took place in a controlled environment to ensure consistency, as shown in Figure 3a. The imaging room was maintained at a constant ambient temperature (approximately 23 °C) and low air current to avoid spurious temperature fluctuations. Before imaging, each subject was asked to rest for a few minutes in the room to acclimate and to remove any jewelry or external items from the hands. The patients sat in front of the prototype device and placed their hands in the field of view of the thermal camera as instructed. The camera-to-hand distance was standardized (around 30 cm) using a fixed mount for the device, and the angle was approximately top-down, capturing the dorsal aspects of the hands (palm-down position). For each subject, eight thermal images were captured—four images per hand. These four images included different orientations or gestures to capture comprehensive thermal information: for example, an image of the dorsum of the hand as shown in Figure 3b, one of the palms, and two lateral or oblique views (such as making a loose fist vs. fingers extended). This approach aimed to capture thermal signatures of all major joints in the hand (fingers, knuckles, wrist), which are common sites of arthritis changes. If a patient had visible inflammation in particular joints (e.g., swollen finger joints in RA or Heberden's nodes in OA), the images would reflect the higher temperature in those areas.

All captured images were labeled according to the subject's diagnosis. Images from healthy individuals were labeled "healthy," images from patients with osteoarthritis were labeled "arthrosis," and images from patients with inflammatory arthritis were labeled "arthritis." The labeled dataset comprised a mix of classes; to ensure balanced learning, classes were balanced either by selective sampling or by augmentation during training. The entire dataset of images was then divided into training, validation, and test sets. We partitioned the data on a per-patient basis—i.e., images from the same patient were kept within the same subset to avoid the classifier learning person-specific cues. Approximately 70% of the patients (and their images) were allocated for model training, 10% for validation during training, and the remaining 20% for hold-out testing (n = 270 images from 20 patients for the test set). This resulted in a test set of images from patients not seen by

the model during training, against which final performance metrics were evaluated. In total, the test set included an equal representation of the three classes (we ensured roughly equal numbers of healthy, arthritis, and arthrosis images for fair evaluation). The primary evaluation metrics were accuracy, confusion matrix, and ROC curves for each class. We also examined class-wise precision, recall (sensitivity), and specificity derived from the confusion matrix. These performance results are presented in Table 2, which is located in the Results section. Figure 4 shows some labeled images from the training set and Figure 5 show the process of data augmentation for one arthritis hand.



(**a**) (**b**)

Figure 3. (**a**) Protocol applied for data collection with real arthritis patients at the UNEVE University Clinic. (**b**) Thermal image of the left hand of an arthritis patient, showing temperature distribution.

Table 2. Classification performance metrics derived from the confusion matrix on the test set (n = 270 images from 20 patients).

Class	Precision	Recall (Sensitivity)	Specificity	F1-Score
Healthy	1	1	1	1
Arthritis	0.41	0.178	0.88	0.25
Arthrosis	0.74	0.744	0.88	0.74
Overall Accuracy	—	—	—	64%

Notes: Metrics were calculated per class using a one-vs.-rest approach. Specificity is computed as the proportion of true negatives (all non-target classes) correctly identified. Precision = TP/(TP + FP); Recall = TP/(TP + FN); F1 = 2 × (Precision × Recall)/(Precision + Recall).

To evaluate the performance of the model, we calculated standard classification metrics including precision, recall, specificity, and F1 score. Precision refers to the proportion of correctly predicted positive observations to the total predicted positives, while recall refers to the ability of the model to detect all relevant cases. These definitions support the interpretation of the confusion matrix and performance indicators reported in Table 2.

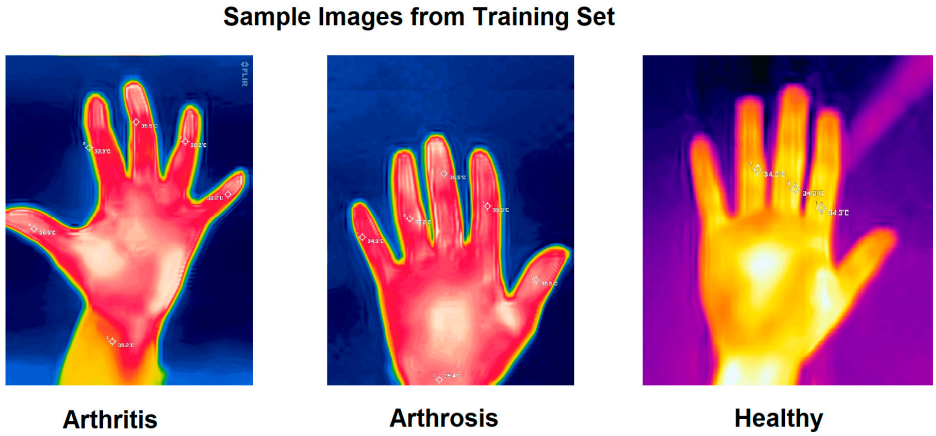


Figure 4. Labeled images from the training set.

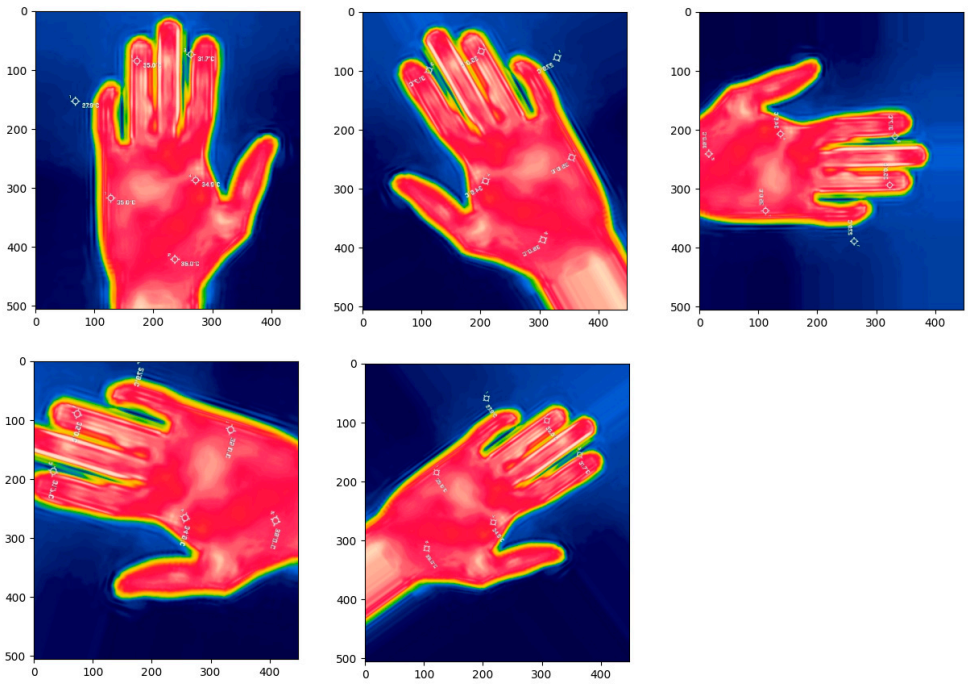


Figure 5. Data augmentation process for an arthritis hand in the training set.

3. Results

The left plot in Figure 6 shows the loss (error) decreasing during training (blue curve) and validation (orange curve), while the right plot shows the accuracy increasing for both training (green curve) and validation (red curve). Over ~200 epochs, the training loss steadily declined, and the training accuracy approached 100%, indicating the model learned the training examples very well. The validation curves followed a similar trajectory, with the validation accuracy reaching around 98% and the validation loss converging to a low value by the end of training. This close tracking of validation performance relative to training performance suggests that no severe overfitting occurred during training—the model was able to generalize well to the validation subset. The final trained model thus achieved high performance on the data it was trained and validated on, motivating the next step of evaluating it on the independent test set of unseen patient images.

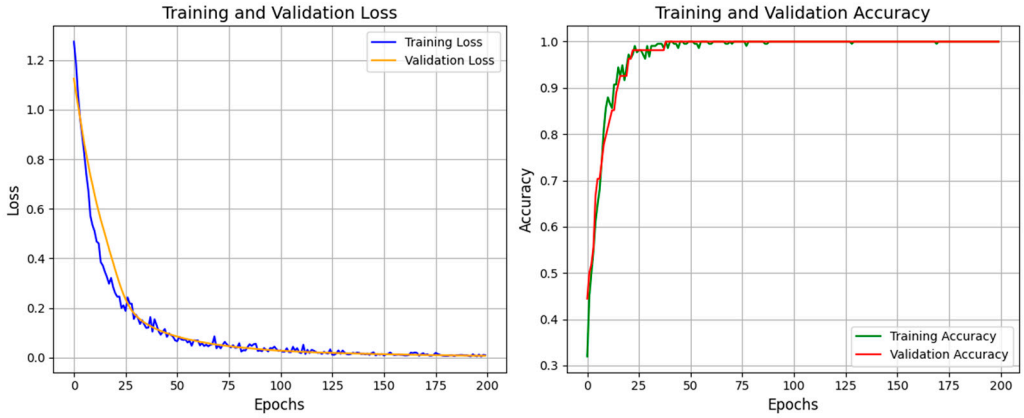


Figure 6. Training and validation performance of the ResNet50 model over the training epochs.

After training the deep learning model, we evaluated its diagnostic performance on the hold-out test dataset consisting of new thermal images from 20 patients (not used in training). The prototype system's predictions on these images were compared to the ground truth diagnoses to compute various metrics. Overall, the system achieved a moderate overall accuracy on the test set. In aggregate, about 64% of the test images were correctly classified into the appropriate category (healthy, arthritis, or arthrosis). While this accuracy is lower than the near-perfect accuracy observed during validation, it reflects the more challenging nature of new patient data and underscores the importance of analyzing performance class by class.

As shown in Table 2, the model achieves perfect precision, recall, and F1 score for the healthy class, indicating that it can reliably identify individuals without joint pathology. However, performance is markedly lower for arthritis, with only 17.8% sensitivity (recall), meaning the model fails to detect most inflammatory arthritis cases. In contrast, the arthrosis class shows moderate performance, with 74.4% sensitivity and 74% precision. The low precision for arthritis (41%) is due to frequent misclassification of arthritis cases as arthrosis, as seen in the confusion matrix. These results highlight the model's strength in ruling out disease (high negative predictive value) but its limited ability to differentiate between arthritis subtypes.

The confusion matrix shown in Figure 7 summarizes the model's predictions across the classes: Arthrosis (Artrosis), Arthritis (Artritis), and Healthy (Sanos). Each row represents the actual true class of the images, and each column represents the class predicted by the model. The diagonal cells (highlighted) indicate correct classifications, while off-diagonal cells indicate misclassifications. As shown, all healthy instances were correctly identified by the model (90/90 healthy images predicted as healthy, with 0 misclassified), demonstrating 100% sensitivity for the Healthy class.

The confusion matrix is based on the hold-out test set, comprising 270 thermal images from 20 patients (90 images per class). Values represent the count of images correctly or incorrectly classified. This implies the system had no false positives for disease among the healthy subjects—a desirable outcome for a screening tool to avoid alarming false alerts. In contrast, the model had difficulty distinguishing between arthrosis and arthritis cases. For images actually belonging to arthrosis patients, 67 were correctly predicted as "Arthrosis," but 23 (approximately 25%) were misclassified as "Arthritis." Similarly, for images from arthritis patients, only 16 were predicted as "Arthritis" (true positive for that class), whereas the majority—74 images—were misclassified as "Arthrosis." In other words, the model tended to predict the degenerative arthrosis class for many of the inflammatory arthritis cases. This resulted in a low sensitivity for the Arthritis class (only ~17.8%, i.e., 16/90 arthritis images were recognized as arthritis), whereas sensitivity for the Arthrosis

class was higher (~74.4%, 67/90). There were no cases of misclassifying a diseased hand as healthy (i.e., zero arthritis/arthrosis images were predicted as healthy), so the model's specificity for detecting "any disease" was excellent; however, its specificity between the two disease types was limited. These results indicate that while the prototype reliably detects the presence of joint abnormalities (separating sick vs. healthy), it struggles in subclassifying the type of pathology from hand thermal patterns alone. The confusion between arthritis and arthrosis suggests overlapping thermal characteristics or insufficient distinctive features captured by the model for these two conditions.

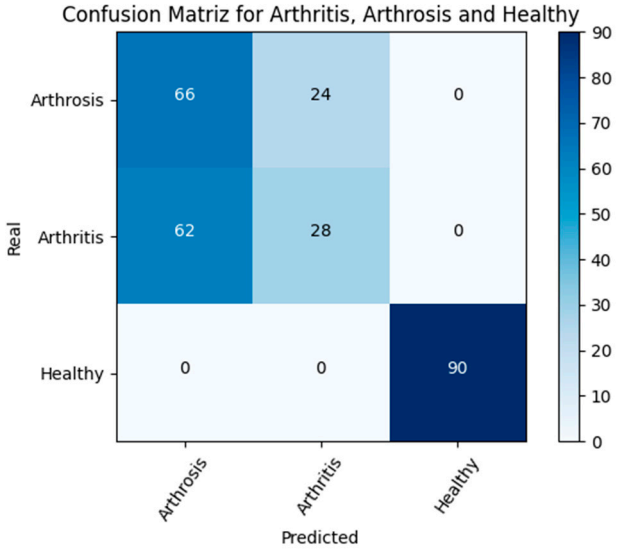


Figure 7. Confusion matrix for the three-class classification on the test dataset (Actual class vs. Predicted class).

To further analyze the classifier's performance, we plotted the ROC curves and calculated the AUC for each class (treating each class in a one-vs.-rest manner). The ROC curve shown in Figure 8 illustrates the trade-off between true positive rate (sensitivity) and false positive rate for different threshold settings of the classifier's output for a given class.

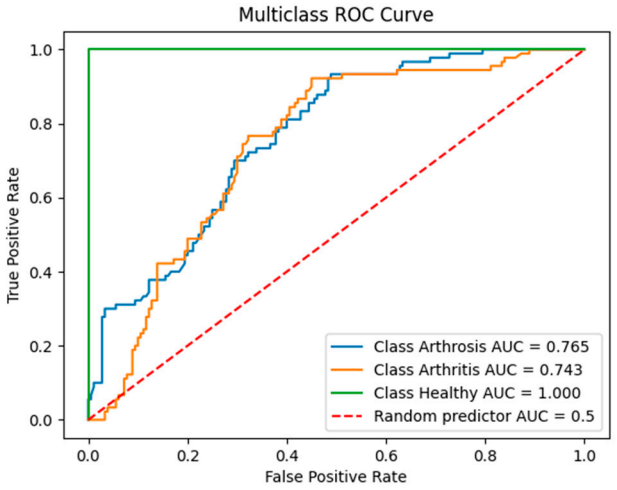


Figure 8. ROC curves for each class (Arthrosis, Arthritis, and Healthy) obtained from the test results of the model.

For each class, the ROC curve is generated by considering that class as "positive" and the other two classes as "negative," and then varying the decision threshold. The green curve corresponds to the Healthy class. It can be seen hugging the top-left corner of

the plot, and its AUC is effectively ~1.00 (100%). This reflects that the model can almost perfectly distinguish healthy hands from those with pathology—indeed, as noted, it made no mistakes in identifying healthy vs. diseased in the test set. The blue curve corresponds to the Arthrosis class, with an AUC around 0.60, and the orange curve corresponds to the Arthritis class, with an AUC around 0.68. These AUC values, being only slightly above 0.5 (the value for random guessing), quantitatively confirm the difficulty the model had in differentiating between arthrosis and arthritis. The Arthritis ROC (orange) is slightly higher than Arthrosis, indicating that the model performs marginally better at identifying arthritis compared to others, rather than vice versa, but both are relatively weak. The curves show that at various threshold choices, the sensitivity and 1-specificity for arthritis/arthrosis trade off modestly above the chance level. In practical terms, the ROC analysis suggests that if the system were used primarily as a binary screener (healthy vs. any joint disease), it would perform extremely well (almost no false negatives or false positives for detecting a problem). However, for triaging between arthritis types, the current model's discriminative ability is limited—an observation consistent with the confusion matrix findings.

Despite the less-than-ideal separation of arthritis subtypes, some qualitative observations were made during analysis of the thermal images. In many arthritis (inflammatory) cases, the thermograms showed diffuse or multiple joint heat patterns (e.g., warmth across several finger joints and wrists), whereas arthrosis (osteoarthritis) cases sometimes showed more localized heat increases over specific joints (such as the base of the thumb or individual finger joints affected by degeneration). The neural network, however, did not fully capitalize on these differences, possibly due to the complexity of patterns or the size of the training data. Nonetheless, the fact that all diseased conditions were distinguished from normal with high confidence is encouraging for the use of thermography in early detection—even if the precise classification between types might require additional refinement.

The Results demonstrate that our thermography-based prototype can successfully perform non-invasive screening of joint health. The system reliably flags individuals with abnormal joint thermal patterns (achieving perfect precision in identifying healthy subjects and high recall for the "any disease" condition). The challenge that emerged is in the finer classification of the specific disease etiology (rheumatoid/inflammatory vs. osteoarthritic). In the following section, we discuss methods and tones of these findings, compare the approach to standard diagnostic methods, and suggest ways to improve the differentiation capability in future work.

To better understand how the ResNet50 model learns and distinguishes thermal patterns associated with each diagnostic category, we analyzed representative input images and their transformations across the three processed color channels used for model inference. Figure 9 illustrates sets of thermographic images of hands, corresponding to patients diagnosed with arthritis, arthrosis, and healthy controls, respectively. For each sample, four visualizations are shown: the original thermal image and three channel variations reflecting the transformed inputs as received by the CNN. These transformations simulate the three-channel (RGB-like) input expected by pre-trained models such as ResNet50 and are derived by applying color mappings or duplicating intensity values with slight alterations.

As shown in the images, cases labeled as arthritis (Figure 9a) exhibit more diffuse and widespread thermal activity, especially around the wrist, metacarpophalangeal joints, and proximal interphalangeal joints. In contrast, arthrosis samples (Figure 9b) tend to present localized high-temperature zones, particularly over distal finger joints or the base of the thumb, consistent with typical osteoarthritic involvement. For healthy subjects (Figure 9c), thermal gradients appear smoother and more uniform, with lower peak temperatures and reduced thermal asymmetry between fingers.

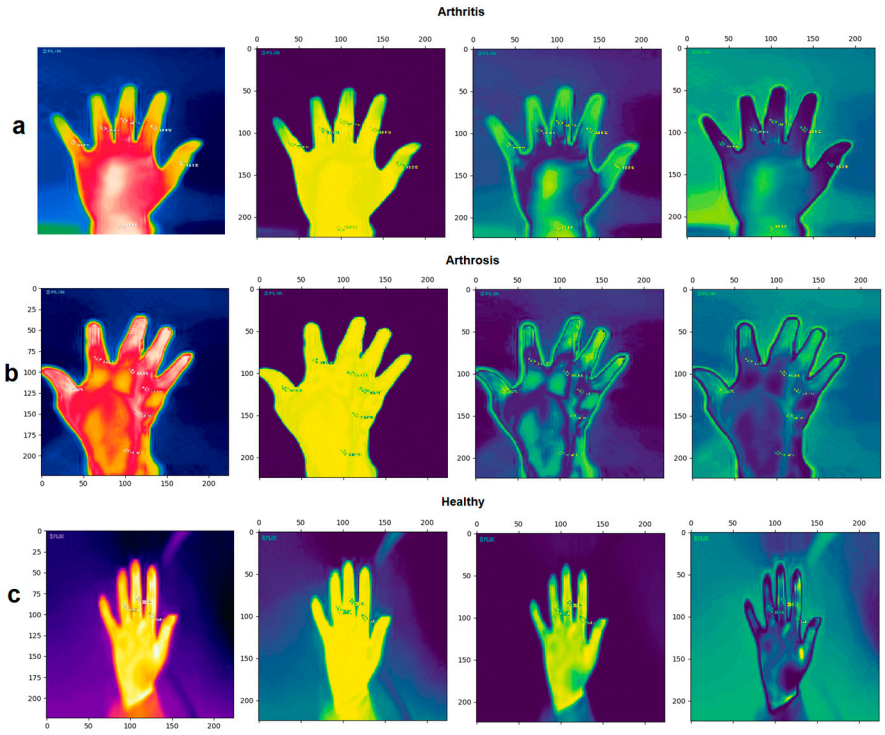


Figure 9. Set of thermographic images of hands corresponding to patients diagnosed with arthritis (**a**), arthrosis (**b**), and healthy (**c**) controls.

The three-channel transformations enable the CNN to extract spatial and textural features from different aspects of the thermal map, which are then processed through convolutional layers to identify class-specific patterns. These visualizations suggest that the model is able to detect and exploit both the distribution and intensity of thermal emissions to infer the underlying pathology. Such channel-wise visualization also supports model transparency and interpretability, helping clinicians and developers understand what the AI is "seeing" during inference and potentially guiding further refinement of preprocessing pipelines to enhance diagnostic accuracy.

4. Discussion

The development and testing of this infrared thermography and deep learning prototype yield valuable insights into its potential and limitations as a clinical tool for early arthritis and arthrosis diagnosis. The results confirmed our primary hypothesis that thermographic patterns of the hand contain discriminative information to detect joint disease [18]. All healthy subjects in the test set were correctly identified by the model, indicating that the absence of significant inflammation or degenerative change in joints manifests as a normal thermal profile that the AI can recognize with near certainty. This is an important finding because a screening device must confidently identify healthy individuals to minimize false alarms. In this regard, the system's performance was excellent—consistent with other studies that have noted clear thermal distinctions between arthritic patients and healthy controls [1,19]. Additionally, ref. [8] reported that a handheld thermal imaging device, combined with machine learning, could distinguish between arthritic patients and healthy ones with over 80% accuracy, highlighting significant differences in temperature and texture features between the groups [9,20,21]. Our work reinforces these observations, as the AI effectively separated "normal" vs. "abnormal" in thermal images of hands.

The more nuanced task of differentiating arthritis vs. arthrosis proved challenging for the prototype. The confusion matrix showed substantial misclassification between these two disease categories. There are several possible explanations for this outcome.

First, overlapping clinical features: Both rheumatoid arthritis and osteoarthritis involve joint inflammation, especially in the hands, and thus both can produce elevated temperatures in affected areas [22,23]. While the underlying causes differ (systemic synovitis in RA vs. local degenerative inflammation in OA), a thermal camera may capture similar hotspots of increased heat in the joints. The thermal patterns might not be distinct enough for a general CNN to separate without additional context. Second, dataset limitations: It is possible that our training dataset did not have sufficient examples or diversity to learn subtle differences. The number of arthritis patients (inflammatory) might have been smaller than that of arthrosis patients, which could bias the model. Moreover, variations in disease severity and distribution (RA can affect wrists and multiple symmetric joints, whereas OA in the hand often affects specific joints like distal interphalangeal joints and the first carpometacarpal joint) could introduce internal class variability that confuses the classifier [24,25]. Specifically, while we balanced the number of images per class using data augmentation, the underlying patient cohort was imbalanced: only 30 patients were diagnosed with inflammatory arthritis (RA or similar), compared to 40 with osteoarthritis (arthrosis), and 30 healthy individuals. Although augmentation helped balance the training data at the image level, the model may still have learned features more representative of the more frequently observed arthrosis cases. This could explain the tendency to misclassify arthritis images as arthrosis—a form of bias stemming from patient-level imbalance. Furthermore, the thermal signatures of early-stage inflammatory arthritis and advanced osteoarthritis can be remarkably similar, both exhibiting localized or diffuse joint hyperthermia due to synovial inflammation [26]. Without access to temporal data (e.g., chronic vs. acute inflammation patterns) or anatomical context (e.g., joint space narrowing on X-ray), the model lacks the discriminative cues that a rheumatologist would use [26,27]. Additionally, the global average pooling layer in ResNet50 aggregates spatial features across the entire hand, potentially diluting subtle differences in thermal distribution (e.g., symmetric polyarticular involvement in RA vs. asymmetric distal joint involvement in OA). Future work will explore region-of-interest (ROI) analysis, attention mechanisms, or hierarchical classification strategies to improve subtype differentiation. Third, spatial resolution and ROI: The hands have many small joints packed closely; our images captured the whole hand, and the ResNet50 processed it holistically. It is possible that a more targeted analysis, focusing on individual joint regions, or a higher-resolution thermal sensor could capture distinguishing details (e.g., RA might cause a more uniform warmth across all knuckles, whereas OA might show isolated spots) [27]. The current model might be averaging these patterns in a way that washes out differences. In future work, techniques such as thermal image segmentation (to isolate each joint or region) and the use of attention mechanisms in the CNN could help the model focus on critical differences.

4.1. Comparison with Conventional Diagnostic Modalities

The concept of using thermography for joint disease must be evaluated in the context of existing diagnostic technologies. The gold-standard methods for diagnosing and assessing arthritis involve a combination of imaging and laboratory tests. For instance, X-rays are routinely used to identify osteoarthritic changes like joint space narrowing, osteophyte formation, and bone erosions in rheumatoid arthritis—but these changes typically appear only after significant disease progression [27,28]. Early-stage arthritis often shows normal X-rays even when the patient has pain or inflammation. Ultrasound and MRI can detect earlier changes: ultrasound can visualize synovial thickening and fluid (synovitis) in RA, and even erosions or osteophytes. MRI can detect bone marrow edema and tiny erosions and is very sensitive to inflammation. However, these tools are expensive and not feasible for mass screening or very frequent monitoring due to cost, required equipment and

expertise, and, in MRI's case, time and patient discomfort. Thermography, by contrast, is fast, non-invasive, and inexpensive. A thermographic scan of both hands takes only a few seconds and does not require specialized personnel to operate (beyond positioning the patient), especially when coupled with an AI that automatically interprets the images. This makes it attractive as an initial screening or triage tool. For example, thermography could be used in primary care or community settings to identify individuals who have abnormal joint heat patterns suggestive of active arthritis, and who can then be referred for definitive evaluation (e.g., by a rheumatologist with ultrasound or lab tests). Our results support this use case: the system would have successfully picked out all patients who had a joint condition (none of the diseased patients were erroneously labeled healthy), meaning its negative predictive value for disease is high. In practical terms, a negative result (classified as healthy) can be trusted to rule out significant arthritis, while a positive result (classified as either arthritis or arthrosis) indicates that the person should undergo further examination. Another important comparison is with clinical examinations by a physician. Doctors assess joint warmth by touch and look for swelling and pain to identify arthritis. This is subjective and can miss subtle inflammation—for instance, a clinician's hand might not detect a mild temperature rise, or warmth might be attributed to ambient conditions. Thermography provides an objective measurement of skin temperature differences, potentially catching early inflammation that is not obvious externally. Indeed, prior research has found thermography sensitive for detecting active synovitis in RA patients even when joints appear clinically normal [24,29]. In our study, even though we did not specifically correlate with clinical exam findings, the very high sensitivity for disease suggests the thermographic camera was picking up on inflammation-related heat that marks disease presence. A key advantage of thermography with AI is its non-invasive and radiation-free nature. Unlike X-rays (which expose patients to ionizing radiation) or blood draws (invasive), an IR scan poses no risk or pain, making it repeatable as often as needed. This is particularly beneficial for monitoring chronic conditions. For example, a patient with rheumatoid arthritis could theoretically use such a device at home or at each clinic visit to monitor joint inflammation over time, with the AI detecting flares by changes in thermal patterns. This could complement clinical scores and help in adjusting treatments earlier. Studies have indeed used thermography for monitoring treatment response, showing a correlation between temperature reductions and effective therapy [1].

4.2. Thermography in the Age of AI—Benefits and Challenges

The integration of AI is a game-changer for medical thermography. Raw thermograms require interpretation. AI algorithms like our ResNet50 model can learn complex thermal signatures of disease that may not be apparent to the naked eye or a simple threshold. As documented in a recent review, machine learning enhances the diagnostic accuracy of thermography. It broadens its applications, leading to systems that can autonomously screen for conditions ranging from breast cancer to diabetic neuropathy by recognizing temperature patterns. Our prototype exemplifies this synergy in the context of joint disease. The key benefits observed include speed and automation—the device produces a result in seconds, which is immediately understandable (e.g., "arthritic changes detected"); objectivity—the AI applies consistent criteria across patients, eliminating inter-observer variability; and cost-effectiveness—the hardware is relatively low-cost and the per-scan cost is negligible, which is advantageous for widespread screening programs. Furthermore, the portability of the unit (small size, potentially battery-powered) means it can be used in varied settings, including rural clinics or bedside, without the need for dedicated imaging facilities.

There are also challenges and limitations to address. One challenge is the standardization of imaging protocol. We maintained a controlled environment in our study, but in real-world usage, ambient temperature, room drafts, or recent patient activity (e.g., coming in from the cold or having exercised the hands) could affect readings. Strict guidelines (acclimatization period, room conditions) and possibly algorithmic normalization will be necessary for consistency. Another issue is the current limited specificity between different disease states, as seen with arthritis vs. arthrosis misclassification. While for a screening tool this may be acceptable (both are "positive" findings needing follow-up), for an automated diagnostic tool to be more clinically useful, improvements are needed. Strategies to improve this could include collecting a much larger dataset with more varied examples of each condition, employing more advanced AI models (for instance, multi-stream networks that consider both temperature distribution and perhaps visible-light images of the hand in tandem), or implementing a hierarchical classification approach (first deciding healthy vs. diseased, then a specialized classifier to distinguish RA vs. OA among the diseased). It may also be beneficial to incorporate clinical metadata into the model (age, symptoms, etc.) since OA and RA often affect different demographics and joint patterns.

Comparatively, while thermography is excellent for early detection of physiological changes, it does not provide anatomical visualization. Thus, it should be seen as complementary to, rather than competitive with, modalities like X-ray or MRI [1]. For instance, if our system flags a patient as having abnormal joint heat (possible arthritis), the next step would be to perform an ultrasound or X-ray to confirm joint damage and make a definitive diagnosis. Thermography with AI can act as an initial filter or a continuous monitor. In resource-limited settings where advanced imaging is not readily available, a thermography-based tool could have a significant impact by identifying patients who most need further care. Its low cost and portability mean it could even be deployed in community health programs to screen older adults for osteoarthritis or at-risk populations for rheumatoid arthritis onset.

As highlighted in other domains, the fundamental advantages of thermography—low-cost, portability, and painlessness—make it particularly valuable for developing communities or widespread screening initiatives. Our study has some limitations to acknowledge. The sample size, while decent for a prototype demonstration, is relatively small for training a deep network and drawing broad conclusions. A hundred patients (with only a subset having arthritis vs. arthrosis) limits the complexity the model can learn. We mitigated this partly with data augmentation, but expanding the dataset is important for future work. Additionally, all images were of hands; thus, the system's applicability to arthritis in other joints (knees, shoulders, etc.) is not tested here. Hands were chosen because they are common sites for both RA and OA, and are easy to image; similar approaches could be applied elsewhere. Another limitation is that we did not compare the thermography results directly with an established metric of disease activity (for example, ultrasound Power Doppler scores or blood inflammation markers). Such a correlation analysis could further validate thermography's effectiveness in quantifying inflammation. The confusion between arthritis types could also partially stem from possible misdiagnosis or overlap in our patient cohort—for instance, some patients classified as "arthrosis" might have had subclinical inflammatory arthritis, or vice versa, which the AI picked up in thermal signals. Rigorous clinical characterization of patients in future studies will help clarify this. Despite these limitations, our findings align with the growing evidence in the literature that thermography augmented by AI is a viable method for early medical diagnosis [1].

In fields like oncology, as mentioned, it is already proving its worth; in musculoskeletal medicine, there is ongoing research echoing our results that thermal imaging can detect joint inflammation effectively [11]. Importantly, thermography is not intended to replace

detailed diagnostic tests, but to serve as an adjunct. In our comparative analysis, each modality has its strengths: X-ray/MRI show structure, ultrasound shows real-time inflammation with high resolution, and thermography shows a quick overview of physiological change. The combination of these, rather than any single one, will likely yield the best diagnostic accuracy. For example, a workflow could be patient reports joint pain → quick thermography scan by nurse/technician → if positive for abnormal heat, refer for ultrasound or rheumatology consult → confirm diagnosis and initiate treatment. This kind of tiered approach could optimize resource use in healthcare. Another dimension is patient comfort and non-invasiveness. Thermography is contactless; this is not only convenient but also eliminates the risk of inducing pain during examination (for instance, pressing on an inflamed joint in a physical exam can be painful, whereas an IR scan does not touch the patient). Especially in chronic conditions requiring frequent monitoring, a painless method encourages patient compliance. During our trial, patients reacted positively to the thermographic procedure—many were curious to see the "heat image" of their hands, which also has an educational aspect (some could visually appreciate which joints were "hotter," correlating with their symptoms).

4.3. Future Perspectives

Building on the current prototype, several improvements are envisioned. On the hardware side, integrating a higher-resolution thermal camera would allow finer thermal details to be captured, potentially improving the classification of different arthritis types. Also, adding a regular RGB camera aligned with the thermal camera could allow hybrid models that use both thermal and visual features (for example, swelling or redness visible in normal light, along with heat in thermal). On the software side, exploring more advanced deep learning architectures (such as attention-based networks or ensemble models) might improve performance. We also consider implementing a two-stage model: one network to detect the presence of pathology (which our results show is very robust), followed by a second network (perhaps requiring more specialized training data) to distinguish arthritis vs. arthrosis. Another promising approach is the use of thermographic radiomics—extracting quantitative features (statistical texture measures, shape of hot areas, etc.) from the thermal images and feeding them into machine learning classifiers. This could complement the CNN's automated feature learning with domain-informed features.

Moreover, broader clinical validation is necessary. A larger multi-center study would help evaluate the system under varied conditions and populations, as well as address potential confounders (like different skin tones, which can slightly affect emissivity, or co-morbid conditions affecting circulation). It would also be valuable to track patients over time: can the thermography prototype detect changes as a patient's condition improves or worsens? If so, it might become a tool not just for diagnosis but also for disease activity monitoring—similar to how disease activity scores or imaging are used in rheumatology to guide treatment adjustments. Our comparative analysis underscores that AI-assisted thermography offers a combination of benefits unmatched by any single conventional modality: it is non-invasive, radiation-free, fast, and low-cost, making it particularly well-suited for early screening and frequent monitoring. The trade-off is that it provides functional rather than structural information and currently lacks specificity between different pathologies. Nonetheless, the ability to catch early signs of disease (inflammation heat) is precisely what is needed for conditions like RA, where early intervention yields the best outcomes. As technology and algorithms advance, it is foreseeable that thermographic devices like the one presented could become common in clinics—much like thermometers or blood pressure cuffs—giving clinicians a quick initial assessment of joint health. Our prototype study is a step in that direction, demonstrating feasibility and guiding future refinements

to fully realize the vision of a smart, infrared-based early warning system for arthritis and arthrosis.

5. Conclusions

In this work, we designed and evaluated a novel prototype system for the early diagnosis of arthritis and arthrosis based on infrared thermography and deep neural networks. The device, built with accessible hardware (Raspberry Pi, thermal camera, touchscreen in a 3D-printed case), exemplifies a low-cost and portable solution for clinical thermal imaging. A Flask-driven application provides a user-friendly interface for image capture and real-time AI-driven diagnosis. Clinically, our pilot study confirmed that thermal patterns of the hands contain salient information to identify joint disease: the ResNet50-based model could clearly discriminate healthy individuals from those with arthritic conditions using only non-contact thermal images. This demonstrates the potential of thermography as an early screening tool for joint pathology—detecting physiological changes (inflammation-related heat) that precede or accompany disease. The system's performance in distinguishing between rheumatoid/inflammatory arthritis and osteoarthrosis was limited, highlighting a key area for improvement.

We provided a thorough comparative discussion, noting that while traditional imaging like X-rays and MRI remain essential for detailed diagnosis, an AI-augmented thermography device offers unique advantages in speed, safety, and cost that can significantly augment current diagnostic pathways. The non-invasive nature of the approach means it can be applied repeatedly with no risk, making it suitable for ongoing monitoring and use in populations (or settings) where conventional imaging is impractical. Future development will focus on increasing the diagnostic specificity of the model and broadening its applicability. This includes expanding the dataset (both in terms of size and diversity of cases), refining the machine learning approach (potentially by using tailored networks or multimodal data fusion), and integrating the prototype into a larger clinical workflow for evaluation. We also envision exploring the use of such thermography systems for other joints (e.g., knees or ankles) and other inflammatory or degenerative conditions. Additionally, rigorous standardization of the imaging process and environment will be pursued to ensure that the system's output is reliable under various conditions, which is critical for real-world deployment.

Integration of infrared thermography with deep learning, as demonstrated by our prototype, represents a promising advancement in medical technology for early disease detection. For arthritis and arthrosis, this approach can facilitate earlier diagnosis and intervention by revealing signs of joint inflammation in a quick and patient-friendly manner. While not a replacement for confirmatory diagnostics, it can serve as a powerful complementary tool—one that leverages the strengths of modern AI to interpret subtle physiological signals. With continued research and validation, AI-powered thermographic screening could become part of the standard arsenal in preventive rheumatology and orthopedics, ultimately contributing to improved patient outcomes through earlier and more accessible detection of joint diseases.

Author Contributions: Conceptualization and methodology, F.-J.A.-C.; clinical validation and authorization, M.-E.H.-L.; social impact analysis, H.-N.L.-F.; formal analysis, L.-M.M.-V.; investigation and rheumatology relation and bioethical committee proposal, J.-L.C.-A.; funding acquisition, F.-J.A.-C.; informed consent design and aproval, A.A.-R.; prototype test and validation, P.R.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institute of Technology of Mexico (TecNM) through the 2025 call for Research and Technological Development Projects, approved with official letter no. M00.2.2/2516/2025.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board as well as Ethics Committee of UNIVERSIDAD ESTATAL DEL VALLE DE ECATEPEC AND INSTIUTO NACIONAL DE REHABILITACIÓN LUIS GUILLERMO IBARRA (protocol code CEI-04-XI-ORD-2024 and approval letter INRLGII 101/25, date of approval 31 July 2025)." for studies involving humans.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data collected in this study will be publicly available in a shared folder on GitHub (https://github.com/CCAITESE/Thermography, accessed on 1 September 2025).

Acknowledgments: This research was performed with the administrative and technical support from Tecnológico de Estudios Superiores de Ecatepec, Universidad Estatal del Valle de Ecatepec, and Instituto Nacional de Rehabilitación Luis Guillermo Ibarra.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Schiavon, G.; Capone, G.; Frize, M.; Zaffagnini, S.; Candrian, C.; Filardo, G. Infrared Thermography for the Evaluation of Inflammatory and Degenerative Joint Diseases: A Systematic Review. *Cartilage* **2021**, *13*, 1790S–1801S. [CrossRef] [PubMed]
2. Aletaha, D.; Aletaha, D.; Aletaha, D.; Smolen, J.S.; Smolen, J.S. The definition and measurement of disease modification in inflammatory rheumatic diseases. *Rheum. Dis. Clin. N. Am.* **2006**, *32*, 9–44. [CrossRef]
3. Gizińska, M.; Gizińska, M.; Rutkowski, R.; Rutkowski, R.; Szymczak-Bartz, L.; Szymczak-Bartz, L.; Szymczak-Bartz, L.; Romanowski, W.; Romanowski, W.; Straburzyńska-Lupa, A.; et al. Thermal imaging for detecting temperature changes within the rheumatoid foot. *J. Therm. Anal. Calorim.* **2020**, *145*, 77–85. [CrossRef]
4. Vicnesh, J.; Salvi, M.; Hagiwara, Y.; Yee, H.Y.; Mir, H.; Barua, P.D.; Chakraborty, S.; Molinari, F.; Rajendra Acharya, U. Application of Infrared Thermography and Artificial Intelligence in Healthcare: A Systematic Review of Over a Decade (2013–2024). *IEEE Access* **2025**, *13*, 5949–5973. [CrossRef]
5. Brioschi, G.C.; Brioschi, M.L.; Dalmaso Neto, C.; O'Young, B. The Socioeconomic Impact of Artificial Intelligence Applications in Diagnostic Medical Thermography: A Comparative Analysis with Mammography in Breast Cancer Detection and Other Diseases Early Detection. In *Artificial Intelligence over Infrared Images for Medical Applications*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2023; Volume 14298, pp. 1–31. [CrossRef]
6. Rakhunde, M.B.; Gotarkar, S.; Choudhari, S.G. Thermography as a Breast Cancer Screening Technique: A Review Article. *Cureus* **2022**, *14*, e31251. [CrossRef]
7. Sims Woodhead, G.; Varrier-Jones, P.C. Investigations on Clinical Thermometry: Continuous and Quasi-Continuous Temperature Records in Man and Animals in Health and Disease. *Lancet* **1916**, *187*, 173–180. [CrossRef]
8. Shaye, K.; Nadav, S.; Keren, N.; Lior, L.; Yael Pri-Paz, B.; Tayer-Shifman, O.; Rotem, S.-H. Accurate Detection of Arthritis Using Hand-held Thermal Imaging and Machine Learning. In Proceedings of the ACR Convergence 2024, Washington, DC, USA, 18 November 2024; p. 76.
9. Devereaux, M.D.; Devereaux, M.D.; Parr, G.; Parr, G.R.; Parr, G.; Thomas, D.P.; Thomas, D.P.; Hazleman, B.L.; Hazleman, B.L.; Hazleman, B.L. Disease activity indexes in rheumatoid arthritis; a prospective, comparative study with thermography. *Ann. Rheum. Dis.* **1985**, *44*, 434–437. [CrossRef]
10. Sudoł-Szopińska, I.; Sudoł-Szopińska, I.; Jans, L.; Jans, L.; Teh, J.; Teh, J. Rheumatoid arthritis: What do MRI and ultrasound show. *J. Ultrason.* **2017**, *17*, 5–16. [CrossRef]
11. Yousefi, B.; Sharifipour, H.M.; Maldague, X.P.V. Embedded Deep Regularized Block HSIC Thermomics for Early Diagnosis of Breast Cancer. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 4504809. [CrossRef]
12. Petrigna, L.; Amato, A.; Roggio, F.; Trovato, B.; Musumeci, G. Thermal threshold for knee osteoarthritis people evaluated with infrared thermography: A scoping review. *J. Therm. Biol.* **2024**, *123*, 103932. [CrossRef] [PubMed]
13. Amisha, F.; Malik, P.; Pathania, M.; Rathaur, V.K. Overview of artificial intelligence in medicine. *J. Family Med. Prim. Care* **2019**, *8*, 2328–2331. [CrossRef] [PubMed]

14. Nowakowski, A.Z.; Kaczmarek, M. Artificial Intelligence in IR Thermal Imaging and Sensing for Medical Applications. *Sensors* **2025**, *25*, 891. [CrossRef] [PubMed]
15. Mishra, S.; Mishra, S.; Mishra, S.; Prakash, A.; Prakash, A.; Roy, S.K.; Roy, S.K.; Sharan, P.; Sharan, P.; Mathur, N.; et al. *Breast Cancer Detection Using Thermal Images and Deep Learning*; IEEE: New York, NY, USA, 2020. [CrossRef]
16. Zhang, Y.; Zhang, Y.; Pan, C.; Pan, C.; Chen, X.; Chen, X.; Wang, F.; Wang, F. Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling. *J. Comput. Sci.* **2018**, *27*, 57–68. [CrossRef]
17. Collins, A.J.; Collins, A.J.; Ring, E.F.J.; Ring, E.F.; Cosh, J.A.; Cosh, J.A.; Bacon, P.A.; Bacon, P.A. Quantitation of thermography in arthritis using multi-isothermal analysis. I. The thermographic index. *Ann. Rheum. Dis.* **1974**, *33*, 113–115. [CrossRef]
18. Li, X.; Li, M.; Yan, P.; Li, G.; Jiang, Y.; Luo, H.; Yin, S. Deep Learning Attention Mechanism in Medical Image Analysis: Basics and Beyonds. *Int. J. Netw. Dyn. Intell.* **2023**, *2*, 93–116. [CrossRef]
19. Aboutalebi, H.; Pavlova, M.; Gunraj, H.; Shafiee, M.J.; Sabri, A.; Alaref, A.; Wong, A. MEDUSA: Multi-scale Encoder-Decoder Self-Attention Deep Neural Network Architecture for Medical Image Analysis. *Front. Med.* **2021**, *8*, 821120. [CrossRef]
20. Li, X.; Li, L.; Jiang, Y.; Wang, H.; Qiao, X.; Feng, T.; Luo, H.; Zhao, Y. Vision-Language Models in medical image analysis: From simple fusion to general large models. *Inf. Fusion.* **2025**, *118*, 102995. [CrossRef]
21. Lin, H.; Xu, C.; Qin, J. Taming Vision-Language Models for Medical Image Analysis: A Comprehensive Review. *arXiv* **2025**, arXiv:2506.18378. [CrossRef]
22. Sultan, R.I.; Zhu, H.; Li, C.; Zhu, D. BiPVL-Seg: Bidirectional Progressive Vision-Language Fusion with Global-Local Alignment for Medical Image Segmentation. *arXiv* **2025**, arXiv:2503.23534.
23. Salisbury, R.S.; Parr, G.; De Silva, M.; Hazleman, B.L.; Page-Thomas, D.P. Heat distribution over normal and abnormal joints: Thermal pattern and quantification. *Ann. Rheum. Dis.* **1983**, *42*, 494–499. [CrossRef]
24. Tan, Y.K.; Hong, C.; Li, H.H.; Allen, J.C.; Thumboo, J. A novel combined thermography and clinical joint assessment approach discriminates ultrasound-detected joint inflammation severity in rheumatoid arthritis at more joint sites compared to thermography alone. *Int. J. Rheum. Dis.* **2022**, *25*, 1344–1347. [CrossRef]
25. Bullock, J.; Bullock, J.; Rizvi, S.A.A.; Rizvi, S.A.A.; Rizvi, S.A.A.; Rizvi, S.A.A.; Rizvi, S.A.; Saleh, A.; Saleh, A.M.; Ahmed, S.; et al. Rheumatoid Arthritis: A Brief Overview of the Treatment. *Med. Princ. Pract.* **2018**, *27*, 501–507. [CrossRef]
26. Mate, G.; Mate, G.S.; Kureshi, A.K.; Kureshi, A.K.; Singh, B.K.; Singh, B.K. An Efficient CNN for Hand X-Ray Classification of Rheumatoid Arthritis. *J. Healthc. Eng.* **2021**, *2021*, 6712785. [CrossRef] [PubMed]
27. Borojević, N.; Borojević, N.; Kolarić, D.; Kolarić, D.; Grazio, S.; Grazio, S.; Grubišić, F.; Grubišić, F.; Antonini, S.; Antonini, S.; et al. Thermography hand temperature distribution in rheumatoid arthritis and osteoarthritis. *Period. Biol.* **2011**, *113*, 445–448.
28. Pauk, J.; Pauk, J.; Ihnatouski, M.; Ihnatouski, M.; Wasilewska, A.; Wasilewska, A. Detection of inflammation from finger temperature profile in rheumatoid arthritis. *Med. Biol. Eng. Comput.* **2019**, *57*, 2629–2639. [CrossRef] [PubMed]
29. Tan, Y.K.; Hong, C.; Li, H.; Allen, J.C.; Thumboo, J. Thermography in rheumatoid arthritis: A comparison with ultrasonography and clinical joint assessment. *Clin. Radiol.* **2020**, *75*, 963.e17–963.e22. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.