rheumato

MDPI

Article

Feature Extraction and Identification of Rheumatoid Nodules Using Advanced Image Processing Techniques

Azmath Mubeen [1,*] and Uma N. Dulhare [2]

[1] Department of Computer Science, University College for Women, Koti, Hyderabad 500095, India
[2] Computer Science & Artificial Intelligence Department, Muffakham Jah College of Engineering & Technology, Hyderabad 500034, India
* Correspondence: azmathmubeen2012@gmail.com or azmathresearch2018@gmail.com

Abstract: Background/Objectives: Accurate detection and classification of nodules in medical images, particularly rheumatoid nodules, are critical due to the varying nature of these nodules, where their specific type is often unknown before analysis. This study addresses the challenges of multi-class prediction in nodule detection, with a specific focus on rheumatoid nodules, by employing a comprehensive approach to feature extraction and classification. We utilized a diverse dataset of nodules, including rheumatoid nodules sourced from the DermNet dataset and local rheumatologists. Method: This study integrates 62 features, combining traditional image characteristics with advanced graph-based features derived from a superpixel graph constructed through Delaunay triangulation. The key steps include image preprocessing with anisotropic diffusion and Retinex enhancement, superpixel segmentation using SLIC, and graph-based feature extraction. Texture analysis was performed using Gray-Level Co-occurrence Matrix (GLCM) metrics, while shape analysis was conducted with Fourier descriptors. Vascular pattern recognition, crucial for identifying rheumatoid nodules, was enhanced using the Frangi filter. A Hybrid CNN–Transformer model was employed for feature fusion, and feature selection and hyperparameter tuning were optimized using Gray Wolf Optimization (GWO) and Particle Swarm Optimization (PSO). Feature importance was assessed using SHAP values. Results: The proposed methodology achieved an accuracy of 85%, with a precision of 0.85, a recall of 0.89, and an F1 measure of 0.87, demonstrating the effectiveness of the approach in detecting and classifying rheumatoid nodules in both binary and multi-class classification scenarios. Conclusions: This study presents a robust tool for the detection and classification of nodules, particularly rheumatoid nodules, in medical imaging, offering significant potential for improving diagnostic accuracy and aiding in the early identification of rheumatoid conditions.

Keywords: rheumatoid nodules; feature extraction; graph-based analysis; multi-class classification; SHAP values

check for updates

Citation: Mubeen, A.; Dulhare, U.N. Feature Extraction and Identification of Rheumatoid Nodules Using Advanced Image Processing Techniques. *Rheumato* **2024**, *4*, 176–192. https://doi.org/10.3390/rheumato4040014

Academic Editor: Giulia Cassone

Received: 28 August 2024
Revised: 10 October 2024
Accepted: 15 October 2024
Published: 24 October 2024

Copyright: © 2024 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/).

1. Introduction

A simple nodule is a small, solid mass or lump of abnormal tissue that can be found in various parts of the body. Rheumatoid nodules, in contrast, are firm, benign subcutaneous lesions commonly associated with rheumatoid arthritis (RA). These nodules typically form on extensor surfaces, such as the elbows, hands, and forearms, and can range in size from small lumps to larger, more prominent growths. Though non-cancerous, the presence of rheumatoid nodules is often indicative of more severe RA and may influence a patient's treatment and prognosis [1].

The formation of these nodules is linked to the systemic inflammation seen in RA, particularly in areas of mechanical stress or trauma. Diagnosing rheumatoid nodules can be challenging, as they may resemble other conditions, such as granulomas, lipomas, or even subcutaneous tumors. Accurate diagnosis is crucial to avoid mismanagement of the underlying RA [2].

The aim of this study is to enhance diagnostic accuracy for rheumatoid nodules by leveraging advanced image processing techniques and feature extraction methods.

1.1. Differential Diagnoses of Rheumatoid Nodules

Rheumatoid nodules can be misdiagnosed as other lesions such as granulomas (caused by inflammatory conditions like sarcoidosis), lipomas (benign fatty lumps) [3], or subcutaneous tumors. Distinguishing between these conditions is crucial, as they require different clinical interventions. A thorough clinical evaluation, combined with advanced imaging techniques and histopathological analysis, is essential for accurate diagnosis.

1.2. Defining Rheumatoid Nodules and Their Differentiating Features

Rheumatoid nodules are unique to RA and possess several differentiating features. They are typically located on extensor surfaces, such as the elbows and hands, in contrast to other benign nodules that can appear in a wider range of locations [4]. Their histopathological structure is characterized by fibrinoid necrosis surrounded by a palisade of histiocytes and chronic inflammatory cells, a structure not seen in other nodules [5]. Additionally, rheumatoid nodules are linked to autoimmune processes and elevated markers of systemic inflammation, distinguishing them from non-inflammatory nodules [6]. Furthermore, rheumatoid nodules tend to grow during active RA, while other benign nodules do not correlate with systemic symptoms [7].

1.3. Advanced Diagnostic Approaches

Current approaches for diagnosing rheumatoid nodules often involve clinical examination, imaging, and histopathological analysis. Magnetic Resonance Imaging (MRI) is the imaging method for evaluating the presence and extent of soft tissue masses which helps in diagnosing the disease easily [8]. However, advancements in image processing and machine learning offer opportunities to improve diagnostic accuracy.

Image Processing and Feature Extraction Techniques

Rheumatoid nodules exhibit a characteristic texture due to their dense, fibrotic nature, which can be effectively captured using texture analysis algorithms. These methods can differentiate the fine textural patterns of rheumatoid nodules from the smoother textures of other benign nodules. The firm, dense, and fibrotic texture of rheumatoid nodules arises from the chronic inflammation associated with rheumatoid arthritis (RA), leading to the formation of fibrous tissues. In image analysis, this translates to high-contrast texture patterns, with distinct variations in surface elevation, density, and irregularity compared to other nodules, which typically exhibit low contrast and a more uniform texture.

Segmentation, color analysis and texture analysis were also used by researchers for feature extraction to generate an accurate mask [9]. Advanced shape analysis techniques are employed to assess the irregular, lobulated contours of rheumatoid nodules. These features differ from the more uniform shapes seen in benign nodules, assisting in their differentiation during image analysis. Additionally, the vascular features associated with rheumatoid nodules, such as increased vascularization, irregular and complicated blood vessels, and hyperemia, serve as critical markers for distinguishing these nodules from non-inflammatory simple nodules. Non-inflammatory benign nodules, such as lipomas or cysts, generally lack these prominent vascular characteristics, which are often avascular in imaging methods like Doppler ultrasound or dynamic contrast-enhanced imaging. The distinctive vascular patterns related to rheumatoid nodules provide a key diagnostic tool for their differentiation from other non-inflammatory lesions.

Finally, the integration of automated histopathological image analysis using machine learning models trained on histopathological images enhances diagnostic precision [10]. These models can accurately identify the characteristic layered structure of rheumatoid nodules, aiding in their differentiation from other lesions. By using these advanced image processing techniques and methods to extract features, this study aims to create a strong

system that accurately identifies rheumatoid nodules. This will help doctors make early diagnoses and improve how they manage and treat patients.

2. Literature Review

Machine learning and deep learning techniques have become valuable tools for analyzing large datasets and uncovering hidden patterns that support decision-making. This capability makes them particularly suitable for being integrated into diagnostic processes for a range of conditions, including rheumatic and neurological disorders, helping physicians deliver more effective healthcare [6,7]. Both supervised and unsupervised machine learning techniques have been employed to study and learn the characteristics of various rheumatic disorders, enabling the prediction of patient conditions for timely intervention. Deep learning techniques have also been used to detect rheumatoid nodules

In this study, we conducted a systematic review of the literature using several key medical and scientific databases, including PubMed, IEEE Xplore, and Scopus, to ensure a comprehensive understanding of the state of the art in machine learning (ML) and deep learning (DL) applications for rheumatic and neurological disorders. Our search spanned articles published between 2000 and 2024. We used a combination of MeSH terms and keywords, such as "machine learning", "deep learning", "rheumatoid nodules", "rheumatic diseases", and "diagnostic tools", to locate relevant studies.

In the research work presented by Ergun et al. [11], the authors considered that the differential diagnosis of benign tumors and tumor-like lesions on the hand and wrist is crucial for determining the appropriate management strategy, whether it be medical, surgical, or observational. MRI is often sufficient for a specific diagnosis when correlated with the patient's history, although diagnostic challenges arise due to the variety of lesions in this region. Common benign masses include giant-cell tumors of the tendon sheath, which typically present as well-demarcated, low-intensity masses in MRI due to hemosiderin-laden cells and dense collagen. Fat-containing masses such as lipomas, macro dystrophia lipomatosa, and neural lipofibromatous hamartomas can be accurately diagnosed through MRI characteristics. Ganglion cysts, often found on the dorsal wrist, present with low signal intensity on T1W images and high intensity on T2W images. Tenosynovitis, an inflammation of the tendon sheath, is identified through MRI by the presence of abnormal fluid thickness and sheath thickening. This comprehensive radiological approach facilitates the accurate differentiation of these lesions, guiding appropriate treatment decisions [11].

Mills et al. [12] present an automated diagnostic tool based on a multilayer neural network designed to detect and classify rheumatic disorders, including rheumatoid arthritis, osteoarthritis, and systemic lupus erythematosus. Trained and tested on a synthetic dataset of 100,000 records, with 70% for training and 30% for testing, the system achieved impressive performance, detecting disorders with an accuracy of 97.48%, a sensitivity of 96.80%, and a specificity of 97.50%. These results indicate that the tool is robust and could be particularly useful in resource-limited settings, where it could facilitate the early detection and treatment of rheumatic diseases by enabling primary healthcare providers to perform tasks typically handled by specialists [12].

Uma et al. [13] focused on the challenging task of classifying nodules, particularly rheumatoid nodules (RNs), from medical images and assessing the associated risks. Leveraging the efficiency of Deep Convolutional Neural Networks (DCNNs) in image analysis, the proposed model employs a dual three-dimensional DCNN design for RN detection and an enhanced Residual Network (Res-Net) for classification. The detection of RNs is initially performed through Pure Deep Convolutional Networks (PDCNNs), followed by feature selection using Particle Swarm Optimization (PSO) and subsequent classification using Res-Net. The integration of deep learning with medical insights aims to minimize errors and reduce false positive (FP) rates, facilitating the early detection of rheumatoid nodules. The study utilizes a dataset collected from various orthopedic hospitals, yielding improved results compared to existing approaches. Additionally, the paper proposes future work

to identify occurrences in the heart and central nervous system, highlighting the ongoing potential for advancements in this area [13].

Muraoka et al. [14] conducted a study investigating the efficacy of magnetic resonance imaging (MRI) texture analysis of the lateral pterygoid muscle in differentiating between rheumatoid arthritis (RA) and osteoarthritis (OA) of the temporomandibular joint (TMJ). The study involved 36 patients, from whom 279 texture features were extracted from MRI images. Using statistical methods, the researchers identified 10 significant texture features that effectively distinguished between RA and OA, achieving high sensitivity (0.78–0.94), specificity (0.89–1.0), and accuracy (0.86–0.92). The findings suggest that MRI texture analysis can provide valuable insights for diagnosing TMJ conditions, potentially improving treatment planning for patients with similar clinical presentations.

Yoo et al. [15] applied the k-means algorithm, an unsupervised learning method, to predict early RA disease risks in patients. Using a clinical dataset of 60 anonymous RA patient records, the data were grouped into four clusters: anti-CCP, Rheumatoid Factor (RF), SJC, and ESR. The k-means model predicted RA disorders based on four assessment parameters: RF > 7, ACCP > 18, SJC > 4, and ESR > 25. The evaluation results indicated that the model accurately predicted RA disease development, with a BSS-to-TSS ratio of 84.1%.

Zhao et al. [16] explored three learning algorithms—logistic regression, support vector machine, and an adaptive boost learning model—for detecting rheumatic diseases using imaging data. The training dataset included 10,058 CT images, with 5000 samples representing sick conditions and 5058 representing normal conditions. The adaptive boost model outperformed the others, achieving a detection accuracy of 90%, compared to 89% for the support vector machine and 79.6% for the logistic regression model.

Gómez Herrero et al. [17] explored the clinical presentation, diagnostic methods, and progression of pulmonary rheumatoid nodules (PRNs) through a detailed analysis of five case studies. Their work emphasizes the importance of recognizing PRNs as a potential complication of rheumatoid arthritis, highlighting the need for accurate imaging and clinical evaluation to differentiate these nodules from other pulmonary conditions. The study's strengths lie in its thorough case analysis, which offers valuable insights into the variability of PRNs and their management strategies. However, a notable limitation is the small sample size, which may impact the generalizability of the findings. Overall, this article contributes significantly to the understanding of PRNs but underscores the necessity for further research to establish more comprehensive diagnostic criteria and treatment protocols. Weaver et al. [18] also emphasized imaging for clinical manifestations in their study.

Gale et al. [4] in their paper provided a comprehensive examination of isolated rheumatoid nodules, emphasizing their clinical presentation, diagnostic challenges, and the importance of a biopsy for accurate diagnosis. The authors presented a detailed case analysis of a 27-year-old male with multiple nonpainful soft tissue masses, highlighting the necessity of differentiating these nodules from other conditions through a biopsy, which is crucial for effective management. The study includes a thorough case documentation and the interdisciplinary approach involving rheumatology and dermatology, promoting collaborative patient care. But the study includes a small sample size, as it is based on a single case, which may affect the generalizability of the findings.

In a book chapter [19] from the book "*Pathology Secrets (Third Edition)*", Damjanov et al. provided a comprehensive overview of the pathology associated with bones and joints, emphasizing the intricate relationship between structural integrity and functional outcomes. The authors systematically discuss various conditions affecting the musculoskeletal system, including inflammatory, degenerative, and neoplastic disorders, while integrating clinical insights with pathological findings. Damjanov's work serves as a valuable resource for understanding the pathology of bones and joints, offering essential knowledge for those in the fields of medicine and allied health [19].

Dang et al. [20] implemented a deep learning model for automatically assigning joint scores to RA patients using X-Ray images and SvH RA image scoring. The model utilized a

13-layer MLP architecture and varied the dimensions of X-Ray images of patients' hands and feet. The evaluation revealed that the deep learning model accurately classified RA patients with an accuracy of 90.8%.

Vodencarevic et al. [21] in their study utilized longitudinal clinical data from the RETRO trial, involving 135 visits from 41 patients. The authors employed four basic machine learning models—logistic regression, k-nearest neighbors, naïve Bayes, and random forests—along with a stacking meta-classifier to predict flare probabilities, achieving an AUROC of 0.81. Key predictors included the percent dose change of biologic drugs, clinical disease activity (DAS-28 ESR), and inflammatory markers. However, the study's limitations include a small sample size and sensitivity to missing data, which may affect the model's generalizability and predictive accuracy.

Zhou et al. [22] in their study employed a machine learning methodology to identify rheumatoid arthritis (RA) patients from primary care electronic health records (EHRs). The researchers linked primary and secondary care EHRs in Wales, UK, using a Random Forest algorithm to select relevant clinical codes, reducing 900 predictors from a total of 43,100 variables to 37 groups.

Gornale et al. [23] used an ensemble of decision trees for detecting OA disorders using knee X-Ray images, with feature extraction from 200 images. The evaluation showed that the model effectively detected OA conditions with an accuracy of 87.92%.

Similarly, Rothschild et al. [24] aims to explore the diverse spectrum of disorders associated with subcutaneous nodules, particularly in the context of rheumatology. The author employed a comprehensive review methodology, synthesizing existing literature to classify and delineate various pathological processes that lead to the formation of these nodules. By addressing the diagnostic implications of these nodules, the study seeks to enhance the understanding of their significance in systemic diseases, thereby aiding clinicians in recognizing underlying conditions.

Growth factors and their relationship was also studied by Badawi et al. [25] who discussed about dysregulation of the growth factor related to molecular pathogenesis of neoplastic and paraneoplastic disease.

Fareez et al. [26] discussed the various challenges in classifying RN's form other forms of lesions specifically in different forms and locations.

Norgeot et al. [27] used a deep learning model to predict rheumatoid arthritis (RA) disorders using EHR data. The study aimed to forecast future outcomes for RA patients, utilizing data sources such as demographics, laboratory results, and prior disease activity measures.

Deane et al. [28] in his study provided a comprehensive overview of rheumatoid arthritis, focusing on its pathogenesis, diagnosis, and treatment strategies. It synthesizes current research and clinical practices to enhance understanding and management of the disease, aiming to improve patient outcome.

Rodolfi et al. [29] provided a comprehensive review of lymphadenopathy as a common clinical finding in rheumatology. The authors utilize a literature synthesis methodology to evaluate the clinical significance of lymphadenopathy in various rheumatologic disorders, including rheumatoid arthritis and systemic lupus erythematosus.

3. Methodology

This methodology integrates mathematical approaches, advanced image processing techniques, and bio-inspired algorithms to accurately identify rheumatoid nodules. The approach leverages feature extraction, classification algorithms, and optimization techniques to enhance diagnostic accuracy.

3.1. Algorithm for Identifying Rheumatoid Nodules in Medical Images

Input: Nodule Image
Output: Classified as Simple Nodule or Rheumatoid Nodule

Step 1: Image Preprocessing

Apply anisotropic diffusion to smooth the image while preserving edges.
Enhance contrast using the Retinex algorithm to improve nodule visibility.
Step 2: Image Segmentation
Segment the image into super pixels using Simple Linear Iterative Clustering (SLIC) to capture localized features.
Step 3: Graph-Based Feature Extraction
Construct a super pixel adjacency graph using Delaunay triangulation.
Extract features such as node degree, edge length, and graph density to distinguish different nodule types.
Step 4: Texture Analysis
Compute texture features like contrast, correlation, energy, and homogeneity using the Gray-Level Co-occurrence Matrix (GLCM).
Step 5: Shape Analysis
Analyze the shape of nodules using Fourier descriptors to characterize the irregular contours of rheumatoid nodules.
Step 6: Vascular Pattern Recognition
Enhance vascular structures using the Frangi filter and extract vascular patterns to differentiate inflammatory from non-inflammatory nodules.
Step 7: Feature Fusion
Integrate global and local features using a Hybrid CNN–Transformer model (ViT for global dependencies, CNN for local features).
Step 8: Classification
Pass the fused features through fully connected layers with softmax activation for multi-class classification or sigmoid activation for binary classification.
Step 9: Bio-Inspired Optimization
Use Gray Wolf Optimization (GWO) for feature selection and Particle Swarm Optimization (PSO) for hyperparameter tuning of classification models.
Step 10: Feature Importance and Interpretability
Apply SHAP values to assess the impact of each feature on the model's predictions, providing both global and local interpretations.

3.2. Data Preprocessing

3.2.1. Image Preprocessing

Apply anisotropic diffusion to smooth the images while preserving edges, followed by contrast enhancement using the Retinex algorithm. This step improves the visibility of nodules and prepares the images for further analysis.

Anisotropic Diffusion

This is used to reduce image noise without blurring edges. The anisotropic diffusion equation is given by

$$\frac{\partial I(x,y,t)}{\partial t} = \nabla\left(c(x,\ y,\ t)\nabla I\left(x,\ y,\ t\right)\right) \tag{1}$$

where

I (x, y, t) is the image intensity at position (x, y) and time t;
∇ denotes the gradient;
c (x, y, t) is the diffusion coefficient, a function of the image gradient.
The diffusion coefficient, c (x, y, t), is typically defined as

$$c\left(x,\ y,\ t\right) = \exp\left(-\left(\frac{|\ \nabla I(x,y,t)\ |}{k}\right)^{2}\right) \tag{2}$$

where k is a contrast parameter controlling the edge preservation.

Retinex Algorithm

Retinex is used for contrast enhancement by estimating the reflectance, R (x, y), of the image.

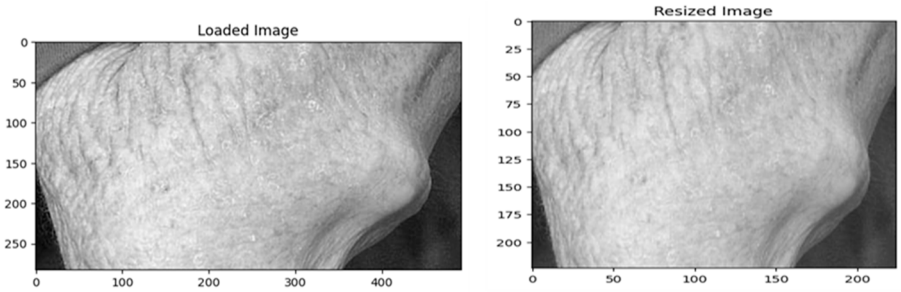
$$R\ (x,\ y) = \frac{I\ (x,\ y)}{L\ (x,\ y)} \tag{3}$$

where

I (x, y) is the observed image;

L (x, y) is the illumination component.

This enhances the contrast by normalizing the image intensity with respect to its illumination. Figure 1a shows the gray-scale image, and Figure 1b gives a resized image. Figure 1a represents the original version of the image, which is larger in both width and height at around 450 × 300 pixels. In Figure 1b, the image has been resized, reducing its dimensions to around 225 × 200 pixels. The general structure and appearance of the lesion remain visible for analysis while optimizing for processing efficiency. The resizing step is used to normalize input images.



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

Figure 1. (**a**) Gray-scale lesion image; (**b**) resized image.

3.3. Superpixel Segmentation

Simple Linear Iterative Clustering (SLIC) is used to segment the images into superpixels. This helps in capturing localized features and reducing the computational complexity. The SLIC algorithm clusters pixels in the xy-color space, minimizing the distance:

$$D = \sqrt{\left(\frac{dc}{m}\right)^2 + ds^2} \tag{4}$$

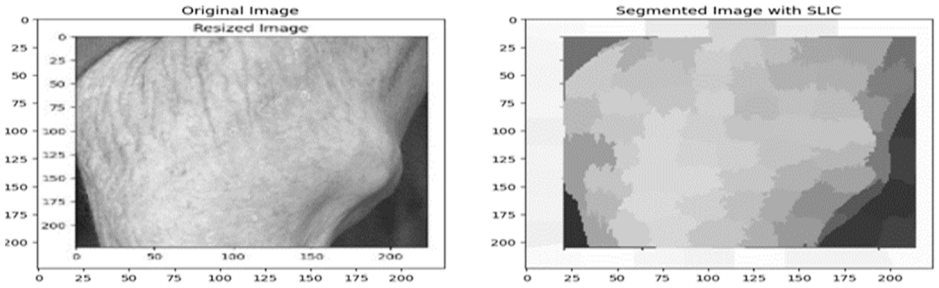
where

dc is the Euclidean distance in the color space;

ds is the Euclidean distance in the spatial space;

m controls the compactness of the superpixels.

Figure 2a shows the output from the previous stage, and Figure 2b shows the image after the segmentation process. Figure 2a shows a resized image of a skin lesion, adjusted for uniformity in size before further image processing. Figure 2b shows the image of the skin lesion segmented using the SLIC algorithm, dividing the lesion into distinct superpixel regions based on similarity in texture and intensity, which aids in more detailed analysis of the lesion's characteristics.



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

Figure 2. (**a**) Resized image; (**b**) segmented image.

3.4. Graph-Based Feature Extraction

A superpixel adjacency graph is constructed using Delaunay triangulation. Features such as the node degree, edge length, and graph density are extracted, which help in distinguishing between different types of nodules.

Delaunay Triangulation

This method creates a graph where nodes represent superpixels and edges represent adjacent superpixels [30]. The feature extraction from the graph is described below.

Node degree:

The number of edges connected to a node, v, is

$$\text{Degree } (v) = \sum_{u \in N(v)} 1 \tag{5}$$

where N (v) is the set of neighboring nodes of v.

Edge length: the Euclidean distance between nodes v and u is

$$\text{Length (e)} = \sqrt{(x_v - x_u)^2 + (y_v - y_u)^2} \tag{6}$$

where (x_v, y_v) and (x_u, y_u) are the coordinates of nodes v and u.

Graph density: this measures the connectivity of the graph:

$$\text{Density} = \frac{2 \mid E \mid}{(\mid V \mid * (\mid V \mid - 1))} \tag{7}$$

where |E| is the number of edges and $\mid$V| is the number of nodes.

Figure 3 displays a Delaunay triangulation graph, where nodes are connected by edges to form a network. The graph density provides a measure of how closely the nodes are interconnected. In this visualization, each red dot represents a node, and the blue lines represent the edges. A higher graph density indicates a more connected network, while a lower density suggests sparser connections.

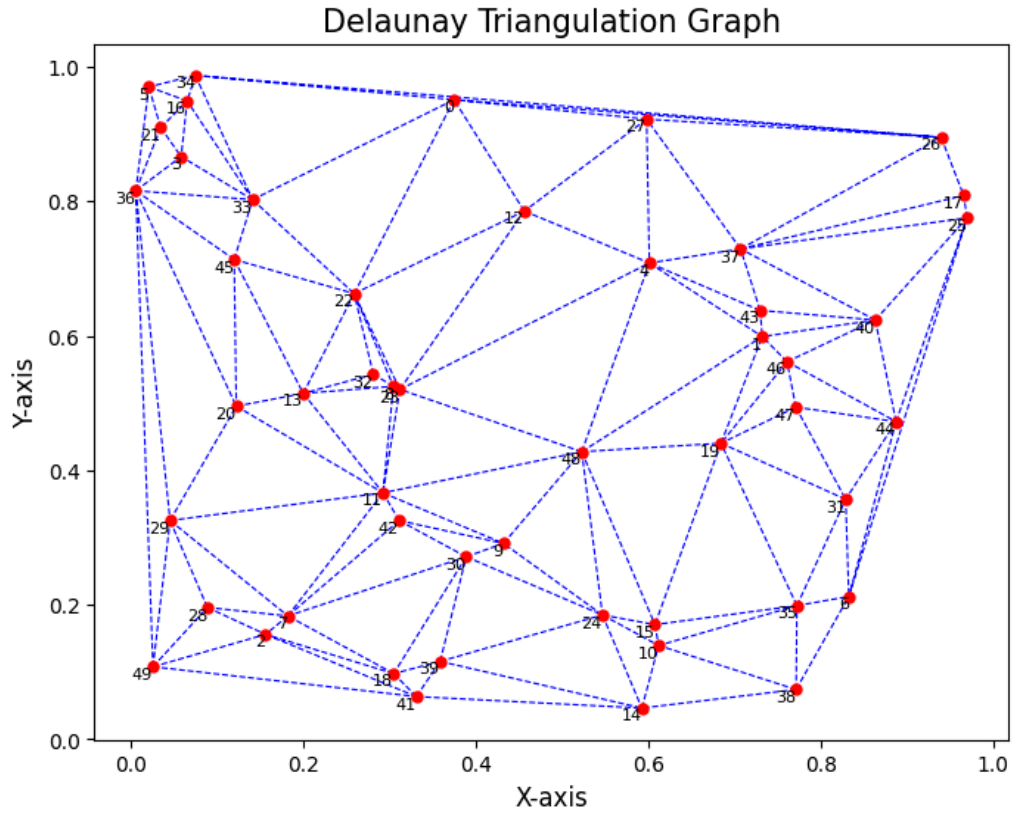


Figure 3. Graph density output.

3.5. Nodule Identification

3.5.1. Texture Analysis

A Gray-Level Co-occurrence Matrix (GLCM) is used to compute texture features such as contrast, correlation, energy, and homogeneity. Rheumatoid nodules typically exhibit higher contrast and lower homogeneity due to their dense, fibrotic structure. The GLCM computes texture features based on the spatial relationship between nodes which is calculated as given in Equation (8).

$$P\ (i,\ j,\ d,\ \theta) = \frac{\sum_{x=1}^{M} \sum_{y=1}^{M} \delta(I(x,y) = i, I(x+dx, y+dy) = j)}{MN} \tag{8}$$

where

P (i, j, d, θ) is the probability of pixel pairs with values i and j at distance d and angle θ;
δ is the Kronecker delta function;
M and N are the dimensions of the image.
Common GLCM features include the following:

- Contrast measures local intensity variation:

$$\text{Contrast} = \sum_{i,\ j} P(i,j)(i-j)^2 \tag{9}$$

- Homogeneity measures the closeness of a distribution:

$$\text{Homogeneity} = \sum_{i,\ j} \frac{P(i,j)}{1+\ |\ i-j\ |} \tag{10}$$

Figure 4 shows the rheumatoid nodule which is identified by the model.

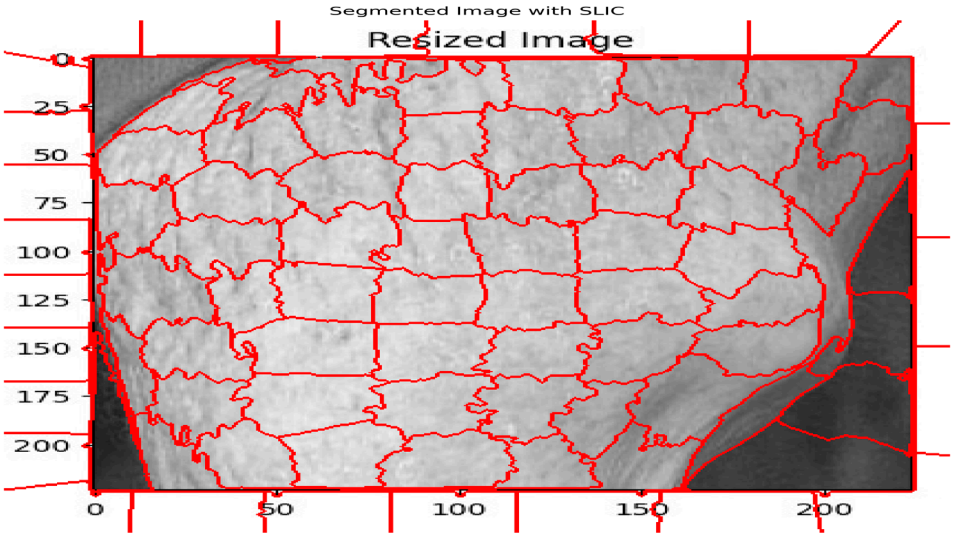


Figure 4. Rheumatoid nodule identified.

3.5.2. Shape Analysis

Fourier descriptors are applied to analyze the shape of the nodules. The irregular, lobulated contours of rheumatoid nodules can be characterized by Fourier coefficients, distinguishing them from the smoother contours of simple nodules.

- These are used to analyze the shape by converting the boundary of the nodule into the frequency domain:

$$Z\,(k) = \sum_{n=0}^{N-1} z(n) exp\left(-\frac{2\pi ikn}{N}\right) \tag{11}$$

where

Z (k) are the Fourier coefficients;
z (n) is the boundary point in the complex plane;
N is the number of boundary points.

3.5.3. Modified Vascular Pattern Recognition

The Frangi filter is used to enhance vascular structures in the images, followed by skeletonization to extract vascular patterns. The tortuosity and branching patterns of the vessels are measured to differentiate inflammatory nodules from non-inflammatory ones. Tubular structures, such as blood vessels, are enhanced by calculating the eigenvalues of the Hessian matrix, H:

$$H = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix} \tag{12}$$

The vesselness function is

$$V\,(x,\,y) = \exp\left(-\frac{\lambda_1^2}{2\alpha^2}\right)\left((1 - exp\left(-\frac{\lambda_1^2}{2\beta^{\,2}}\right)\right) \tag{13}$$

where λ_1 and λ_2 are the eigenvalues of the Hessian matrix, and α and β are parameters controlling the sensitivity to noise and vesselness.

3.6. Classification

3.6.1. Hybrid CNN–Transformer Model–ViT and CNN Integration

A hybrid model combining Vision Transformer (ViT) and a CNN is used for classification. The ViT captures global dependencies in the image, while the CNN extracts local features.

$$\text{HviT} = \text{TransformerEncoder (P(E))} \tag{14}$$

$$\text{H}_{\text{CNN}} = \text{CNN (I}_{\text{norm}}\text{)} \tag{15}$$

$$\text{H}_{\text{fusion}} = \text{Concatenate (H}_{\text{ViT}}\text{, H}_{\text{CNN}}\text{)} \tag{16}$$

where

P (E) represents the positional encoding of image patches;
H_{ViT} and H_{CNN} are the respective feature maps;
H_{fusion} is the combined feature representation.

3.6.2. Final Classification

The fused features are passed through fully connected layers, followed by softmax activation for multi-class classification or sigmoid activation for binary classification.

$$\hat{y} = \text{Softmax } (W \times H_{\text{fusion}} + b) \tag{17}$$

where

W and b are learnable parameters of the fully connected layer.

3.7. Bio-Inspired Optimization Methods

3.7.1. Gray Wolf Optimization (GWO)

GWO is applied for feature selection and parameter tuning of the classification models. GWO mimics the social hierarchy and hunting behavior of gray wolves to optimize the selection of the most significant features, improving model performance. The objective is to maximize accuracy and the F1-score while minimizing the number of features, leading to a more interpretable and efficient model.

The final position update is

$$X\,(t+1) = \frac{X_\alpha\,(t) + X_\beta\,(t) + X_\delta\,(t)}{3} \tag{18}$$

where

X (t + 1) is the updated position vector;
$X_\alpha(t)$, $X_\beta(t)$, and $X_\delta(t)$ are the positions of the alpha, beta, and delta wolves.

3.7.2. Particle Swarm Optimization (PSO)

PSO is used to optimize the hyperparameters of the ViT and CNN. PSO simulates the social behavior of birds flocking to find the optimal set of parameters that enhance model performance. PSO can be summarized in the following steps:

Step 1: Initialization:

- Initialize a population (swarm) of particles, where each particle represents a potential solution in the search space.
- Each particle i has a position vector, x_i (t), and a velocity vector, v_i (t), at time step t.
- Randomly initialize the position, x_i (0), and velocity, v_i (0), of each particle.
- Each particle also remembers its personal best position, p_i (t), (i.e., the best position it has encountered so far) and has access to the global best position, g (t), (i.e., the best position found by any particle in the swarm).

Step 2: Velocity Update:

The velocity of each particle is updated based on its current velocity, its personal best position, and the global best position. The update rule is given by

$$v_i\ (t + 1) = w\ v_i\ (t) + c_1 r_1\ (p_i(t) - x_i\ (t)) + c_2 r_2\ (g(t) - x_i\ (t)) \quad (19)$$

where

w is the inertia weight, controlling the impact of the previous velocity.

c_1 and c_2 are acceleration coefficients that weigh the relative pull of the personal best and global best positions.

r_1 and r_2 are random numbers uniformly distributed between 0 and 1, adding stochasticity to the particle's movement.

Step 3: Position Update:

The position of each particle is updated using the newly computed velocity:

$$x_i\ (t + 1) = x_i\ (t) + v_i\ (t + 1) \quad (20)$$

This equation updates the particle's position in the search space based on its velocity.

Step 4. Personal and Global Best Update:

Update the personal best position, p_i (t), if the new position, x_i (t + 1), is better (based on the fitness function).

Update the global best position, g(t), if any particle's new position, x_i (t + 1), is better than the current global best.

Step 5. Stopping Criterion:

The process is repeated iteratively until a stopping criterion is met, such as a maximum number of iterations or convergence to a solution within a specified tolerance.

3.8. Feature Importance and Interpretability

3.8.1. SHAP Values

SHAP (Shapley additive explanations) values are implemented to assess the impact of each feature on the model's predictions. SHAP values provide a unified measure of feature importance, helping to understand the contribution of each feature to the classification of nodules.

Deep SHAP values: Deep SHAP values are computed to understand the contribution of each input feature to the model's predictions.

$$\phi_j\ = \frac{\sum_{S \subseteq N\{j\}} |S|!\ x\ (|N| - |S| - 1)}{|N|!}\ [f(SU\ \{j\}) - f(S)] \quad (21)$$

where

f(S) represents the output of the deep learning model when features, S, are present.

ϕj is the SHAP value for feature j. It represents the contribution of feature j to the difference between the actual prediction and the average prediction over all possible feature sets.

N is the set of all features in the model. If the model has n features, then N = {1, 2, . . ., n}.

S ⊆ N {j} represents all possible subsets, S, of the feature set, N, that do not include feature j. The sum is taken over all such subsets, S.

|S| is the cardinality (number of elements) of the subset S. It represents the size of the subset of features considered. This is the total number of features in the model.

|N|! represents the factorial of the total number of features, |N|. It is used to normalize the contribution of each subset size.

|S|! is the factorial of the number of elements in the subset S. It accounts for the permutations of the subset S in the overall calculation.

$(|N| - |S| - 1)!$ represents the factorial of the number of elements in the complement of S excluding feature j. It accounts for the remaining features outside of subset S and feature j.

f(S) is the output of the model when only the features in subset S are present. Essentially, it represents the model's prediction based on the features in S.

f(S∪{j}) is the output of the model when both the features in subset S and feature j are present. This term measures the model's prediction when feature j is added to the subset S.

Local Interpretation

For a single prediction, the SHAP value, ϕj, for each feature j represents how much that feature contributes to the deviation of the prediction from the average prediction. The local interpretation is given by the sum of SHAP values for all features, which reconstructs the model's output for a specific instance:

$$f(x) = \phi_0 + \sum_{j=1}^{M} \phi j \quad (22)$$

where

f(x) is the prediction for the specific instance x.

ϕ_0 is the average model output over the training data (i.e., the base value).

ϕ_j is the SHAP value for feature j for the instance x, indicating its contribution to the prediction.

Global Interpretation

To obtain a global interpretation, SHAP values are aggregated across all instances in the dataset to determine the overall importance of each feature. The global importance of a feature j is usually calculated by taking the average absolute SHAP value across all predictions:

$$\text{Global Importance of Feature j} = \frac{1}{n}\sum_{i=1}^{n}\left|\phi_j^{(i)}\right| \quad (23)$$

where

$\phi_j^{(i)}$ is the SHAP value of feature j for the i-th instance.

N is the total number of instances in the dataset.

In the present work, local SHAP values are used for an individual prediction, as SHAP values can reveal which features (e.g., texture, shape, vascular pattern) contribute the most to classifying a specific nodule as a rheumatoid nodule versus a simple nodule. And global SHAP values average the SHAP values across all instances, which means you can determine which features are generally most important in differentiating between rheumatoid and simple nodules across the entire dataset.

4. Results

Assessment of Model Performance

Accuracy is the ratio of correctly predicted instances (both positive and negative) out of all instances in the dataset. It reflects the overall effectiveness of the model in classifying instances correctly.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (24)$$

where

TP (true positive) signifies the count of accurately predicted positive instances.

TN (true negative) signifies the count of properly predicted negative instances.

FP (false positive) denotes the instances of negative classification errors as positives.

FN (false negative) denotes the instances of positive classification errors as negatives.

Precision

Precision measures the accuracy of positive predictions. It is particularly important in medical diagnosis, where false positives can lead to unnecessary treatments. Precision ensures that when the model predicts a positive case, it is likely to be correct.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} \ + \ \text{False Positive}} \tag{25}$$

Recall

Recall (or sensitivity) measures the model's ability to correctly identify all positive cases. In medical diagnosis, recall is critical for detecting as many true positive cases as possible, reducing the chances of missed diagnoses (false negatives).

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} \ + \ \text{False Negative}} \tag{26}$$

F1 Measure

$$\text{F1} = \frac{2x \ Precision \ x \ Recall}{Precision + Recall} \tag{27}$$

Figure 5 presents the confusion matrix, which visually compares the predicted and actual values. The matrix demonstrates the model's effectiveness in classifying the data accurately, providing insight into the distribution of errors (false positives and false negatives).

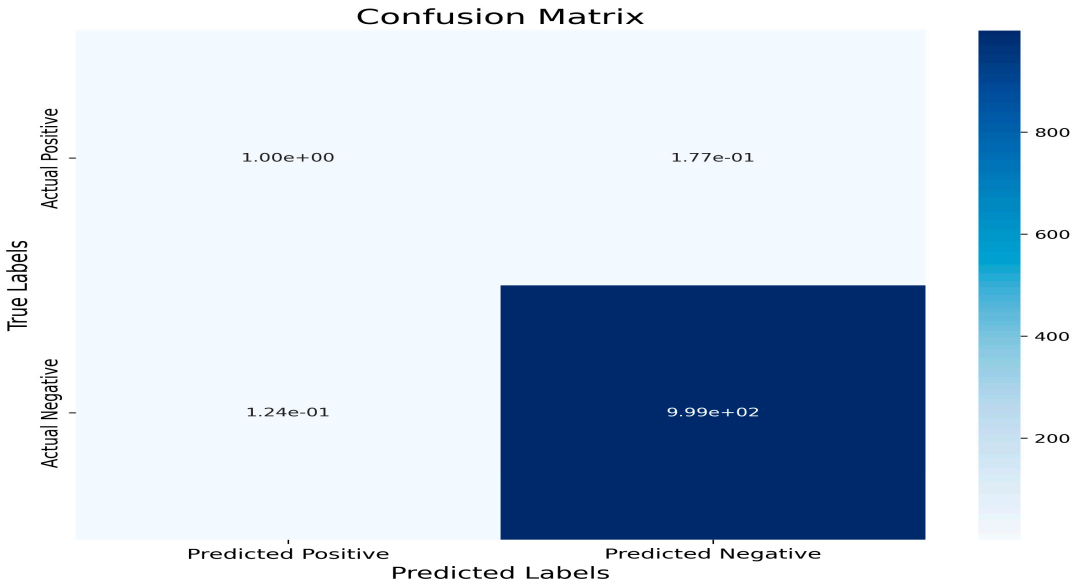


Figure 5. Confusion matrix.

5. Discussion

This research work aims to enhance the diagnostic accuracy of rheumatoid nodules by employing advanced image processing techniques and feature extraction methods. Rheumatoid nodules, characterized by their location on extensor surfaces, distinct histopathological features, and association with autoimmune activity, can be misdiagnosed as other lesions. Table 1 compares the proposed model's performance with existing models for various rheumatic conditions. Previous research emphasizes the importance of these characteristics, along with their correlation to active rheumatoid arthritis. By integrating techniques like texture analysis, shape assessment, and vascular pattern recognition, this study offers significant advancements in diagnosing rheumatoid nodules. However,

limitations such as a small sample size and the focus on technological aspects without fully addressing clinical implications must be acknowledged. Overall, this research highlights the potential of these methods to improve early diagnosis and patient management, ultimately enhancing outcomes for those with rheumatoid arthritis.

Table 1. Comparison of the model with existing models.

Study	Condition	Accuracy	Precision	Recall	F1 Score	Notes
[12]	Rheumatic Disorders	97.48	92.3	96.80%	Not Specified	Multilayer neural network trained on synthetic dataset.
[13]	Rheumatoid Nodules (RNs)	98.80%	96.39%	95.28%	92.14	Dual three-dimensional DCNN with PSO for feature selection and Res-Net classification.
[16]	Rheumatoid Arthritis (RA)	82.7	87.6	85.4	83.3	k-means clustering for predicting early RA disease risks.
[2]	Rheumatic Diseases	90	89.90	87.7	86.7	Adaptive boost model performed best among logistic regression, SVM, and adaptive boost.
[17]	Systemic Sclerosis (SS)	88.7	91.09	92%	88	Random forest algorithm on EHRs with high sensitivity and F1 score.
[20]	Rheumatoid Arthritis (RA)	91%	87.90	89.7	90.7	Deep learning model predicting RA.
[24]	Rheumatoid Arthritis (RA)	92.29%	85.6% (PPV)	86.2	82.4	Ensemble method with high PPV and specificity, using large datasets.
Proposed Model	Rheumatoid Nodules	85%	85	89	0.87	Anisotropic diffusion, superpixel segmentation, graph-based features, texture analysis, Fourier descriptors, vascular pattern recognition, feature fusion with CNN–Transformer model, and bio-inspired optimization (GWO and PSO).

6. Conclusions

This study demonstrates the potential of advanced machine learning techniques for detecting and classifying rheumatoid nodules. By combining traditional image features, graph-based features, and the Hybrid CNN–Transformer model, we achieved high performance in both binary and multi-class classification tasks. The model's ability to integrate various features enhances diagnostic precision, providing valuable tools for early diagnosis in rheumatoid arthritis and related conditions. These techniques offer promising improvements for patient management and clinical outcomes.

Author Contributions: A.M. and U.N.D. conceptualized the study. A.M. was responsible for the methodology and data acquisition, while both A.M. and U.N.D. validated the data and conducted the formal analyses. The original draft of the manuscript was prepared by A.M., with the final review and editing completed by both A.M. and U.N.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding or grant.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable. As only images from online repository were taken.

Data Availability Statement: The data is available online.

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

References

1. Kim, S.; Lee, J.Y.; Lee, J.W. Atypical rheumatoid nodules: A possible precursor to a rheumatoid variant in a rheumatoid-factor-negative patient. *Hand* **2009**, *4*, 62–65. [CrossRef] [PubMed]
2. Guo, Q.; Wang, Y.; Xu, D.; Nossent, J.; Pavlos, N.J.; Xu, J. Rheumatoid arthritis: Pathological mechanisms and modern pharmacologic therapies. *Bone Res.* **2018**, *6*, 15. [CrossRef] [PubMed]
3. Available online: https://my.clevelandclinic.org/health/diseases/15008-lipomas (accessed on 27 August 2024).
4. Gale, M.; Gilbert, E.; Blumenthal, D. Isolated rheumatoid nodules: A diagnostic dilemma. *Case Rep. Med.* **2015**, *2015*, 352352. [CrossRef] [PubMed]
5. Nemakayala, D.R.; Surmachevska, N.; Vaqar, S.; Ramphul, K. Caplan Syndrome. In *StatPearls [Internet]*; StatPearls Publishing: Treasure Island, FL, USA, 2024. Available online: https://www.ncbi.nlm.nih.gov/books/NBK499886/ (accessed on 27 August 2024).
6. Shaw, M.; Collins, B.F.; Ho, L.A.; Raghu, G. Rheumatoid arthritis-associated lung disease. *Eur. Respir. Rev.* **2015**, *24*, 1–16. [CrossRef] [PubMed]
7. Mușetescu, A.E.; Gherghina, F.L.; Florescu, L.M.; Streba, L.; Ciurea, P.L.; Florescu, A.; Gheonea, I.A. Computer-Aided Diagnosis of Pulmonary Nodules in Rheumatoid Arthritis. *Life* **2022**, *12*, 1935. [CrossRef] [PubMed]
8. Peh, W.C.G.; Truong, N.P.; Totty, W.G.; Gilula, L.A. Magnetic resonance imaging of benign soft tissue masses of the hand and wrist. *Clin. Radiol.* **1995**, *50*, 519–525. [CrossRef]
9. Oukil, S.; Kasmi, R.; Mokrani, K.; García-Zapirain, B. Automatic segmentation and melanoma detection based on color and texture features in dermoscopic images. *Skin Res. Technol.* **2021**, *28*, 203–211. [CrossRef]
10. Shah, K.K.; Pritt, B.S.; Alexander, M.P. Histopathologic. *J. Clin. Tuberc. Other Mycobact. Dis.* **2017**, *7*, 1–12. [CrossRef]
11. Ergun, T.; Lakadamyali, H.; Derincek, A.; Tarhan, N.C.; Ozturk, A. Magnetic Resonance Imaging in the Visualization of Benign Tumors and Tumor-like Lesions of Hand and Wrist. *Curr. Probl. Diagn. Radiol.* **2010**, *39*, 1–16. [CrossRef]
12. Mills, G.A.; Dey, D.; Kassim, M.; Yiwere, A.; Broni, K. Diagnostic Tool for Early Detection of Rheumatic Disorders Using Machine Learning Algorithm and Predictive Models. *Biomed. Inform.* **2024**, *4*, 1174–1201. [CrossRef]
13. Dulhare, U.N.; Mubeen, A. Detection and Classification of Rheumatoid Nodule using Deep Learning Models. *Procedia Comput. Sci.* **2023**, *218*, 2401–2410. [CrossRef]
14. Muraoka, H.; Kaneda, T.; Hirahara, N.; Ito, K.; Okada, S.; Kondo, T. Efficacy of magnetic resonance imaging texture features of the lateral pterygoid muscle in distinguishing rheumatoid arthritis and osteoarthritis of the temporomandibular joint. *Dento Maxillo Facial Radiol.* **2023**, *52*, 20220321. [CrossRef] [PubMed]
15. Yoo, J.; Lim, M.K.; Ihm, C.; Choi, E.S.; Kang, M.S. A study on prediction of rheumatoid arthritis using machine learning. *Int. J. Appl. Eng. Res.* **2017**, *12*, 9858–9862.
16. Zhao, J.; Li, L.; Li, J.; Zhang, L. Application of artificial intelligence in rheumatic disease: A bibliometric analysis. *Clin. Exp. Med.* **2024**, *24*, 196. [CrossRef]
17. Gómez Herrero, H.; Arraiza Sarasa, M.; Rubio Marco, I.; García de Eulate Martín-Moro, I. Pulmonary Rheumatoid Nodules: Presentation, Methods, Diagnosis and Progression in Reference to 5 Cases. Available online: https://www.ncbi.nlm.nih.gov/pubmed/352984 (accessed on 10 October 2024).
18. Weaver, J.S.; Vina, E.R.; Munk, P.L.; Klauser, A.S.; Elifritz, J.M.; Taljanovic, M.S. Gouty Arthropathy: Review of Clinical Manifestations and Treatment, with Emphasis on Imaging. *J. Clin. Med.* **2021**, *11*, 166. [CrossRef]
19. Damjanov, I. Bones and Joints. In *Pathology Secrets*; Mosby: St. Louis, MO, USA, 2009; pp. 409–433. [CrossRef]
20. Dang, S.D.H.; Allison, L. Using deep learning to assign rheumatoid arthritis scores. In Proceedings of the 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science, Las Vegas, NV, USA, 11–13 August 2020; pp. 399–402.
21. Vodencarevic, A.; Tascilar, K.; Hartmann, F.; Reiser, M.; Hueber, A.J.; Haschka, J.; Bayat, S.; Meinderink, T.; Knitza, J.; Mendez, L.; et al. Advanced machine learning for predicting individual risk of flares in rheumatoid arthritis patients tapering biologic drugs. *Arthritis Res. Ther.* **2021**, *23*, 67. [CrossRef]
22. Zhou, S.M.; Fernandez-Gutierrez, F.; Kennedy, J.; Cooksey, R.; Atkinson, M.; Denaxas, S.; Siebert, S.; Dixon, W.G.; O'Neill, T.W.; Choy, E.; et al. Defining disease phenotypes in primary care electronic health records by a machine learning approach: A case study in identifying rheumatoid arthritis. *PLoS ONE* **2016**, *11*, e0154515. [CrossRef]
23. Gornale, S.S.; Patravali, P.U.; Manza, R.R. Detection of osteoarthritis using knee X-Ray image analyses: A machine vision-based approach. *Int. J. Comput. Appl.* **2016**, *145*, 20–26.
24. Rothschild, B. Subcutaneous Nodules as Manifestations of Systemic Disease. *Rheumato* **2024**, *4*, 75–87. [CrossRef]
25. Badawi, R.A.; Birns, J.; Watson, T.; Kalra, L. Growth factors and their relationship to neoplastic and paraneoplastic disease. *Eur. J. Intern. Med.* **2005**, *16*, 83–94. [CrossRef]
26. Fareez, F.; Moodley, J.; Popovic, S.; Lu, J.-Q. Rheumatoid nodules: A narrative review of histopathological progression and diagnostic consideration. *Clin. Rheumatol.* **2023**, *42*, 1753–1765. [CrossRef] [PubMed]
27. Norgeot, B.; Glicksberg, B.S.; Trupin, L.; Lituiev, D.; Gianfrancesco, M.; Oskotsky, B.; Schmajuk, G.; Yazdany, J.; Butte, A.J. Assessment of a deep learning model based on electronic health record data to forecast clinical outcomes in patients with rheumatoid arthritis. *JAMA Netw. Open* **2019**, *2*, e190606. [CrossRef] [PubMed]

28. Deane, K.D.; Aletaha, D.; Bathon, J.M.; Holers, V.M.; Kolfenbach, J.R.; St Clair, E.W. Rheumatoid Arthritis. In *Clinician's Pearls & Myths in Rheumatology*; Stone, J.H., Ed.; Springer: Cham, Switzerland, 2023; pp. 1–23. [CrossRef]
29. Rodolfi, S.; Della-Torre, E.; Bongiovanni, L.; Mehta, P.; Fajgenbaum, D.C.; Selmi, C. Lymphadenopathy in the rheumatology practice: A pragmatic approach. *Rheumatology* **2024**, *63*, 1484–1493. [CrossRef]
30. Sunarya, C.A.; Siswanto, J.V.; Cam, G.S.; Kurniadi, F.I. Skin cancer classification using Delaunay triangulation and graph convolutional network. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*, 68–85. [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.