


Review

Rheumatoid Arthritis Diagnosis: Deep Learning vs. Humane

George P. Avramidis ¹, Maria P. Avramidou ² and George A. Papakostas ^{1,*} 

¹ MLV Research Group, Department of Computer Science, International Hellenic University, 65404 Kavala, Greece; geavoam@cs.ihu.gr

² Consultant in Rheumatology, Schmerzklinik Basel, 4010 Basel, Switzerland; avramidmar@gmail.com

* Correspondence: gpapak@cs.ihu.gr; Tel.: +30-2510-462321

Abstract: Rheumatoid arthritis (RA) is a systemic autoimmune disease that preferably affects small joints. As the well-timed diagnosis of the disease is essential for the treatment of the patient, several works have been conducted in the field of deep learning to develop fast and accurate automatic methods for RA diagnosis. These works mainly focus on medical images as they use X-ray and ultrasound images as input for their models. In this study, we review the conducted works and compare the methods that use deep learning with the procedure that is commonly followed by a medical doctor for the RA diagnosis. The results show that 93% of the works use only image modalities as input for the models as distinct from the medical procedure where more patient medical data are taken into account. Moreover, only 15% of the works use direct explainability methods, meaning that the efforts for solving the trustworthiness issue of deep learning models were limited. In this context, this work reveals the gap between the deep learning approaches and the medical doctors' practices traditionally applied and brings to light the weaknesses of the current deep learning technology to be integrated into a trustworthy context inside the existed medical infrastructures.

Keywords: deep learning; rheumatoid arthritis (RA); trustworthiness; explainable AI; artificial intelligence; medical imaging; computer-aided diagnosis



Citation: Avramidis, G.P.;

Avramidou, M.P.; Papakostas, G.A.

Rheumatoid Arthritis Diagnosis:

Deep Learning vs. Humane. *Appl.*

Sci. **2022**, *12*, 10. [https://doi.org/](https://doi.org/10.3390/app12010010)

10.3390/app12010010

Academic Editors: Augusto Ferrante,
Mingcong Deng and Mihaiela Iliescu

Received: 15 November 2021

Accepted: 19 December 2021

Published: 21 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

1.1. Definition and Epidemiology of RA

Rheumatoid arthritis (RA) is a systemic autoimmune disease that preferably affects small joints. RA can also cause many extra-articular manifestations such as pericarditis, pulmonary fibrosis, and peripheral neuropathy, etc. Most often, the diagnosis is made when patients with pain and swelling in the peripheral joints, as well as joint stiffness in the morning, seek medical help. Pain due to RA is typically worse in the morning and the evening and improves during the day. Awakening in the night because of pain in the joints has also been described as a common symptom of RA, as in most inflammatory rheumatic diseases. RA is a “multicausal” disease that most likely results from a combination of genetic predisposition and various environmental and lifestyle factors. Articular and systemic manifestations can lead to poor long-term outcomes such as disability and death [1].

The worldwide prevalence of RA has been estimated as 0.24% based upon the Global Burden of Disease 2010 Study and was approximately two times higher in women [2]. Certain population groups are much more affected than others. This is why the prevalence of RA can reach up to 1% of the general population in some countries. The lifetime risk of developing RA is 3.6% in women and 1.7% in men [3]. The prevalence of RA has remained stable from 1990 to 2010. The annual incidence of RA in the United States and northern European countries is estimated to be approximately 40 per 100,000 persons [4,5]. Globally, of the 291 conditions studied, RA was ranked as the 42nd highest contributor to global disability. Disability was measured in years lived with disability (YLDs) and represented 0.49% of total YLDs. YLDs for RA increased over the years (from 48/100,000 in 1990 to 55/100,000 in 2010). This was due to the population increase worldwide over these years and due to the aging of the population. YLDs were higher in females in comparison to

males [2]. The prevalence of RA is more intense in the regions of central and north Europe, while it is more suppressed in countries near the equator. There are also countries with no prevalence data (e.g., Australia).

Disability-adjusted life years (DALYs) because of RA increased by 44% for all ages over the years 1990–2010. In 2010, RA was ranked 74th in global DALYs, accounting for 0.19% of total DALYs [2]. The age-standardized death rate from RA in 2010 was approximately 0.8 per 100,000, but there is considerable uncertainty regarding this matter [2]. RA's financial burden on society is substantial. The financial burden of RA is a summative effect of RA treatment itself (doctor visits, medication including DMARDs = disease modifying antirheumatic drugs, biologicals, physiotherapy, ergotherapy, psychotherapy if needed, etc.) and the costs of sickness absence. According to a study published in 2009, societal costs of RA in the US are USD 19.3 billion and USD 39.2 billion (in 2005 dollars) without and with intangible costs, respectively [6]. Since 2009, there have been many more biologicals on the market as a therapeutic option for RA, which are more selective and therefore more expensive. The annual treatment costs of specific immunosuppressive drugs used against RA (most of them are also used against other inflammatory rheumatic diseases such as axial spondyloarthritis or psoriatic arthritis) are presented in Figure 1 [7].

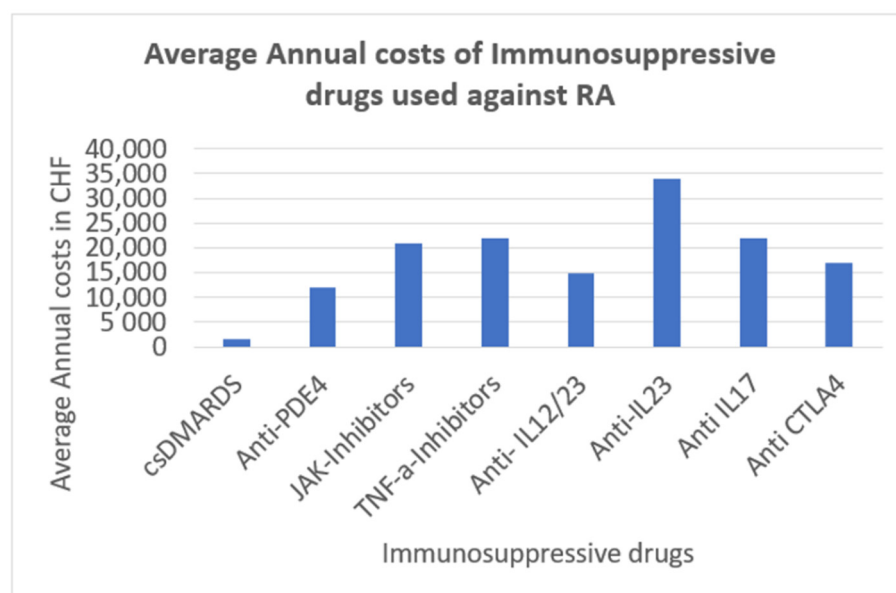


Figure 1. Average annual costs of immunosuppressive drugs against RA [7].

Hence, the medical treatment of the patients is costly, and the degradation of the joints is non-irreversible. Taking into account the growing pain of the patient as well as the deterioration of the quality of life of those people suffering from RA, the prospect of a well-timed diagnosis of RA disease is considered to be quite important.

1.2. Deep Learning on Medical Imaging

The big success that deep learning models have shown during the recent years in image recognition tasks has forced researchers to look for solutions in image classification regarding the medical imaging field. The development of more robust deep learning models coupled with the recent trend of digitalization of the medical records as it is noticed in [8] can help the evaluation of the usage of deep learning in the specific scientific field. As deep learning is heavily connected with image recognition and classification tasks, it seems to be a suitable tool for computer vision tasks on medical images [9]. The main medical image types include X-ray, ultrasound (US), computed tomography (CT) and magnetic-resonance imaging (MRI). It must be noted that historically, the algorithms moved from manual feature extraction methods to feature learning methods, i.e., convolution neural networks (CNN). The digitalization of the medical data that led to the creation of big labeled data

sets, along with the technological progress in the field of computers (graphics processing unit—GPU, tensor processing unit—TPU) have pushed the usage of deep learning models in many medical image analysis tasks combined with continuous improvements regarding the accuracy and the training time of the models.

1.3. Deep Learning on RA

In rheumatology, and more specifically, in the scope of RA, several deep learning researchers over the past years have presented their work as they tried to achieve a method for automatic classification of medical images for RA diagnosis. According to [10], it is very difficult for a doctor not majored in RA to evaluate precisely the condition of RA as its diagnosis is affiliated with implicit knowledge. The authors in [11] also agree that physicians have to rely on a manual and subjective examination of radiographs. In [12], it is noted that the manual assignment of the numerous (86) joints score following the Sharp/van der Heijde—SvH [13] score system is time consuming, expensive by means of human effort and sometimes inaccurate.

As deep learning models mostly need input images, in most cases, radiographs (X-rays) of hands were used in order to automatically assess the SvH score [13] for each image. In some studies [14–16], the technology of Ultrasound imaging was used as well as the corresponding OMERACT-EULAR Synovitis Scoring (OESS) system [17]. Regarding the use of radiographs, it is the most widely used medical modality to diagnose the condition of RA and monitor its activity, as it has low cost and it is easily accessible. In general, regarding the usage of radiographs, most models try to predict if the condition is present, and if so, they try to predict the two types of joint damage: narrowing and erosion. In the case of models that use ultrasound images as input, the model searches for the presence of synovitis and predicts its score according to the OESS system.

1.4. The Trustworthiness Issue

Generally, it is believed that a deep learning model can be identified as being similar to a black box, regarding the way it comes to a result. This implies that there is no precise knowledge of the mechanisms that take place during the model's operation, so it is not perfectly clear why the model came to a specific decision. This situation may not bother a user that employs a deep learning model for object detection tasks or even for face recognition tasks; however, it may not be totally acceptable when it comes to medical and health issues. In this field of applications, the user may be a physician or a patient himself and has the need to understand the outcome of the model to the fullest so that he can justify this decision or accept it easier. Moreover, this is the reason that deep learning models have not been established in clinical use even if they outperform the human experts on some medical diagnostic tasks. In [18], there is a reference to the methods that researchers use in order to be able to explain the decision of a model when it comes to medical diagnostic tasks. On the other hand, in [19] is noticed that even if the progress in medical image processing is huge, some approaches produce erroneous results because of the neglect of prior knowledge. This knowledge may derive from physician expertise and may include additional medical data and medical exams. Hence, the question that has to be answered is whether, e.g., a medical image is sufficient for a diagnosis or there should be a way that this additional knowledge should be taken into consideration so that the model's outcome is explainable in a better way. Regarding the medical field of rheumatology, in this paper we compare the physician method and the deep learning algorithms that have been developed in order to counter the RA disease. The objective of this comparison is to verify if and to what extent the developed deep learning models follow the same method as a physician does for the RA disease confrontation. If there is an extended matching, it can be said that the physician method is followed by the models so the models can be trusted more easily. However, if the models take into account only a part of the physician method or the method that they use is not completely explained, then it can be considered that explainability issues arise. In such a case, the model's outcome cannot be entirely accepted as the model

itself cannot be trusted. In [20], it is suggested that the explainability is too demanding as a requirement but trust in experts is mostly based on their ability to produce certain results and to give reason for their actions.

2. RA Diagnosis Methods: Modernism vs. Traditionalism

2.1. Deep Learning Methods

In this review, we proceed with reference to the state-of-the-art methods that utilize algorithms from the field of deep learning in order to counter the RA disease. To the best of our knowledge, the 14 works that are reviewed in our work constitute the main efforts of the researchers in order to combine the deep learning method and medical data for diagnosing RA disease.

The authors in [11] present a multi-task deep learning model that can learn to detect joints on X-ray images and concurrently diagnose two kinds of joint damage, narrowing damage and erosion damage. Moreover, they propose an alternation for label smoothing. Their way combines cues from classification and regression into a single loss. This way achieves a reduction of 5% regarding the relative error compared to other standard functions. In order to approach the standard metric SvH [13] for hand and foot images, they execute segmentation and classification tasks at the same time. For the training procedure, they utilize four images per patient, with all the scores for narrowing and erosion damage, annotated for each image. In order to make the procedure more robust, they use access to annotations of the center of all joints for better training signals.

In [10], there is a proposal for a method based on deep learning that realizes at the same time recognition and classification of the RA using X-ray images. As it is difficult for a not experienced RA doctor to evaluate the state of the condition, as this diagnosis relies on implicit knowledge, the authors believe that there is a need for the implementation of a system that can evaluate the RA condition automatically. Furthermore, they developed their system in a way so it can be improved by evaluations and modification of the doctor who performs the diagnosis. Their system consists of four specific procedures. At first, there is the generation of the training data by the user. The user in this state detects the position of the joints and measures the size of the image and reads the SvH [13] score that is written on the X-ray image. Therefore, there is a five-dimensional (position x , y , width, height, score) data record that trains their model. At the next step, when there is an input of a new image, the trained model predicts the position of the joints and the damage score (DS) as well. This info is drawn on the image. Next, the medical doctor has the task of checking the result. At this stage, the outcome can be modified (score, position). Finally, this checked data can be fed into the training procedure of the model in order to make the system more robust.

A deep learning model for calculating the radiographic finger joint SvH score in RA using X-ray images was developed in [21]. The model executes two tasks in two steps: the joint detection step and the joint evaluation step. Regarding the first step, a classifier was trained to detect the finger joints using Haar-like features. For the second step, meaning the joint space narrowing (JSN) score and the erosion score assignment for each detected joint, a CNN model was used. This model consisted of seven layers (two convolution layers, two pooling layers and three fully connected layers). Data augmentation (horizontal flipping, rotation) took place during both steps to increase the training data sets. In training sets, the images were manually clipped and scored for JSN and bone erosion by clinicians. The performance of the model was examined using a test dataset by comparing the score assigned by the clinicians and the score assigned by the model.

In [15], they used grayscale ultrasound images as an input for the DesnseNet-121 model in order to automatically classify the conditions of RA. To standardize the use of the ultrasound, the OMERACT-EULER Synovitis scoring (OESS) [17] was used. The dataset of the ultrasound images was created from hospital data after the approval of the corresponding patients. According to the OESS guidelines, the joints that were scanned were the radiocarpal-intercarpal joint and radioulnar joint of the wrist, the proximal interphalangeal

and the metacarpophalangeal joints. The images were then annotated by physicians with experience in ultrasound imaging. The physicians also marked the area of synovial proliferation. As the medical dataset was limited, data augmentation was used in order to create a relatively unbiased training set. The authors focus on two scenarios. The first scenario is associated with the presence of synovial proliferation and the second one is with the medical status of the patient (healthy or diseased). There were three groups of experiments performed with different inputs for the model. The variation was the different sizes of the ROI (region of interest) and the presence or not of a pre-segment mask. Furthermore, the authors insist that transfer learning could be used due to the limited quantity of medical data. It should be noticed that, for better visualization and consequently better explainability of the model outcome, heatmaps were used by means of class activation mapping.

In the work described in [12], the authors present a deep learning model for automatic assignment of joint scores and overall scores for RA patients using X-ray images. They used the SvH [13] RA image scoring. Similar to other researchers, they insist that the manual approach of assigning the SvH score is expensive by means of human time and effort and is sometimes inaccurate and subjective. According to their method, there are two objectives. The first one is associated with the prediction of the narrowing and erosion scores for each joint, trying to maintain high accuracy. The second objective of their method is associated with the prediction of an overall RA score for each patient. As they used a dataset with X-ray images from hands and feet for each patient with varying dimensions, they resized the images to a standard dimension set. A data augmentation method was also used because of the small number of patients. After the augmentation, they combined the four images per patient (left/right hand/foot) so that a new larger image is created for each patient. These images were fed into a deep learning model with 13 layers of depth, achieving high accuracy.

The authors of [22] try to classify RA by using deep learning models to analyze texture changes in different stages of the disease. They use the Deep Texture Encoding Network (Deep-TEN) and residual network-50 (ResNet-50) in order to predict the probability of RA. They use radiographs in order to assess the bone structural and textural changes, which indicate the progress of RA. The method of fractal analysis was used in order to determine bone texture characteristics from the radiograph images. To focus on a specific region of the images, they trained a curve-graph convolutional network (GCN). By using this model, they achieved a fully automated segmentation method of the second, third and fourth metacarpal bone regions. The segmented images were augmented. The authors based their research on the Deep-TEN model, as they insist that such a model can learn the essential features that are necessary to fit and identify the region of interest in an image regarding a specific problem. The Deep-TEN model that was used is a model specialized for texture analysis and includes a novel encoding layer, which is a point of difference in comparison to other CNN models such as ResNet. In that way, it achieves good performance in texture recognition tasks.

Other authors in [14] engaged with the usage of CNN models for OESS [16] (Doppler mode in US—DUS) system assessment on ultra-sound images for diagnosis and monitoring of patients with RA. They used two state-of-the-art CNN models (VGG16, Inception V3) for two tasks. The first model was used for binary image classification as healthy or diseased (0–1), and the second model was used for OESS score assessment (0–3). The results of the two models were compared with the results of a rheumatologist physician. They created four DUS image sets for each OESS score category. The VGG-16 model was used in order to classify the images of RA joint disease activity as healthy (DUS scores 0 and 1) or diseased (DUS scores 2 and 3). The Inception-V3 model, a more sophisticated model, has a modular architecture as it consists of several modules, the so-called inception modules. Each one of these modules extracts information from the input image following different depths, i.e., different resolutions. The information is then combined in a mixed layer. The first mixed layers contain more generic information in comparison with the latest mixed layers. Then they used the information from the mixed layer with the higher classification accuracy in

order to ensemble a classification method where 10 classifiers were trained but with slightly different training parameter settings. The results show that CNN models can be used for DUS image OESS score assessment, as they achieve high accuracy.

Regarding the work of [23], the authors try to use plain hand radiographs in order to create a method that automatically diagnoses rheumatoid arthritis and monitors its activity. They created an image dataset from radiographs coming from clinic files. From raw images, they kept only the right hands, maintaining the same aspect ratio from original images. They also used data augmentation to avoid overfitting and to achieve better overall performance. During the online data augmentation, random vertical and horizontal translation for several pixels and rotation of the image was used. The CNN model that was used by the authors had six groups of convolution layers. Each group consisted of one convolution layer, a batch normalization layer and a ReLU layer. The first five groups also include a max-pool layer. In the end, there is a fully connected layer and a softmax layer.

According to another method described in [24], the authors propose a novel method for automatic detection of bone erosion on hand radiographs. First, their method performs a segmentation of the image in order to extract the region of interest (ROI) that contains the detailed phalanges regions. This selection is performed using the MSGVF Snakes method. Then, as the regions are selected, the method continues by using a deep neural network classifier in order to identify if there is bone erosion in these regions or not. Regarding the creation of the ROI, the proposed method removes at first the soft tissue of the radiograph image using greyscale morphological operations. After that, during the initial segmentation, the phalangeal region is extracted by estimating the joint position of each phalange from the bone region. The bone regions are located using contour tracing. At this stage, the selected regions contain noise. For this reason, the final detailed phalangeal region is created using the MSGVF Snakes method. After this final segmentation, the method can continue with the use of a DCNN so that the model can learn to distinguish if there is bone erosion in the specific region or not. It is noted that from the final 40×40 -pixel segmented image, only the phalange region is loaded into the DCNN. As it is generally difficult to collect medical images for model training, the transfer learning method has been used on several occasions, so it was used by the authors. They used a pre-trained model that was trained using 1000 categories of general images.

The model proposed in [25], the FingerNet model, is an approach for finger detection from hand radiograph images using CNN in order to assist physicians in diagnosing diseases such as rheumatoid arthritis. Hand radiography is the simplest way for disease diagnosis and demands minimal exposure to radiation. Their method needs a little user intervention, and it consists of three different stages: the pre-processing (PP) stage, the finger extraction (FE) stage and the joint detection (JD) stage. During the first stage, the model creates the mask of the hand by executing segmentation in the original radiographic image. In the second stage, the model extracts five separate fingers from the previously created hand mask image. At the final stage, the model detects three points for each finger by the use of CNN architecture and a signal processing-based joint break detection. The CNN method that is used is based on the LeNet-5 model, while the joint break method is a method during which the finger image is scanned from the end of the tip to its base in order to find peak locations where the intensity of the finger image changes dramatically. This advanced method outperformed the AdaBoost model.

Another two-staged model, which is proposed in [26], combines the use of the object detection method and convolution neural networks, which can predict the joint level narrowing and erosion SvH scores [13], as well as the overall RA damage, from patients' radiographs. At the first stage, the model performs object detection using the RetinaNet object recognition models. These models were trained in order to detect finger and wrist joints from radiograph images. During the second stage, the model with the use of CNNs with attention predicts the joint-wise narrowing and erosion SvH scores and the overall RA damage applying the joints that were extracted from the previous stage. The attention mechanism that is used helps the model to focus on the salient regions of the radiograph

images so that the damage predictions are more accurate. Furthermore, the visualization that is added on top of the radiograph images helps with the explainability of the model predictions.

In [27], they use a deep learning model to predict the RA state of the patient at his next clinic visit. As the records of the patients are digitally stored and accessed by EHR (electronic health record) platforms such as university hospitals (UN) and public safety-net hospitals (SNH), it is relatively easy to create datasets with the corresponding data. These data can be used in order to train the deep learning model so that it can forecast the clinical disease activity index (CDAI) score of the patient at his/her next visit. The author selected the ESR and CRP level, the prior CDAI score, DMARDs, oral and injected glucocorticoids, autoantibodies, and finally, the demographic data of the patients as variables for predicting the disease state. Their research showed that only the previous CDAI score was not enough for predicting the next score, and a combination of variables such as laboratory values, medication and the history of disease activity should be used in order to obtain a correct prediction for the next visit. The outcome of their work shows that deep learning models can be used on EHR data for the accurate prediction of the RA disease condition.

Rohrbach et al. [28] followed the Ratingen [13] score system for annotating the X-ray images that fed their model. They used only the left hand radiograph to extract only the joint regions. Thus, ten new images showing the ten joint regions were created from one X-ray image. These images were next rated by expert raters, and then they were used to train the model. The authors used a VGG16-based deep learning model and applied transfer learning. They used the pre-trained VGG16 model as the core, and they replaced some of the fully connected layers. After the replacement, they had to fine-tune the entire model, so that it could adapt to the new domain. However, they conclude that a whole new training of the model from scratch could be the best solution as the datasets are sufficient. Because of the six categories of the Ratingen score system and the non-equal distribution of the data in the datasets, besides the global accuracy of the model, the metrics of balance accuracy and the ± 1 balanced accuracy were introduced. The authors also experimented with a weighted cross-entropy because of the pronounced imbalances in the Ratingen scores in their data, which shows better results. The authors conclude that such a system can predict the Ratingen score as well as a trained expert, but it takes only some milliseconds for the outcome compared to the minutes that an expert has to spend.

An algorithm for automatic segmentation of ultrasound images is proposed in [16]. The segmentation is necessary in order to provide prominence to the different anatomical regions such as bones, skin and joints that are shown in the ultrasound image. These regions are significant for efficient discrimination and alternation of the disease condition. The first step was associated with the pre-processing of the original image so that the noise is eliminated. At the second step, the separation of the skin region is conducted as boundary regions and edges are determined using the Canny edge detection technique. In order for the model to be able to detect the bone region of the image, at the next step, the authors are focused on the intensity variation in the image, as the signal strength of the bone region is represented by high-intensity pixel. Next, the joint region detection is performed by measuring the distance between the two bone regions. The metric that was used was the Euclidean distance. At the fifth step, the model continues with the synovial region detection. This region can be defined by the synovial fluid extension. Hence, the segmentation of this region was performed using an active contour technique. Finally, these segmented regions are fed into the CNN model for classification into four categories. The proposed model achieved high accuracy greater than 95%.

In Figures 2 and 3, we present a general block diagram for the training and the testing process of the reviewed models, respectively. In Figure 4, we present the block diagram that describes the work of [10], in which the part of the physician is essential.

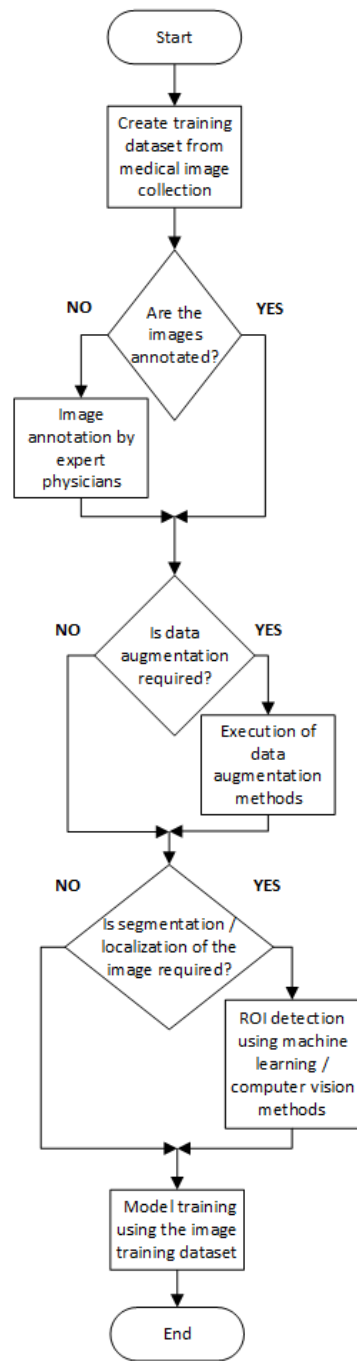


Figure 2. General block diagram for training CNN models.

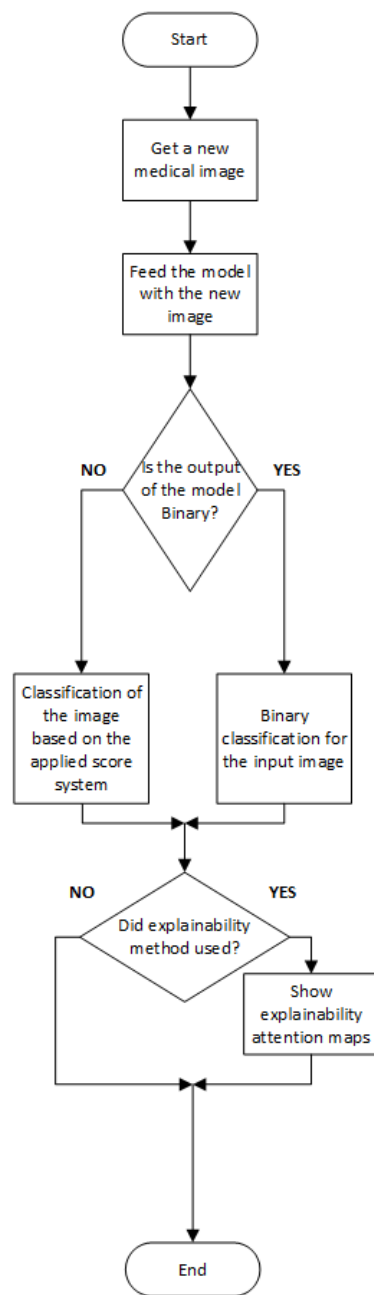


Figure 3. General block diagram for testing the CNN models.

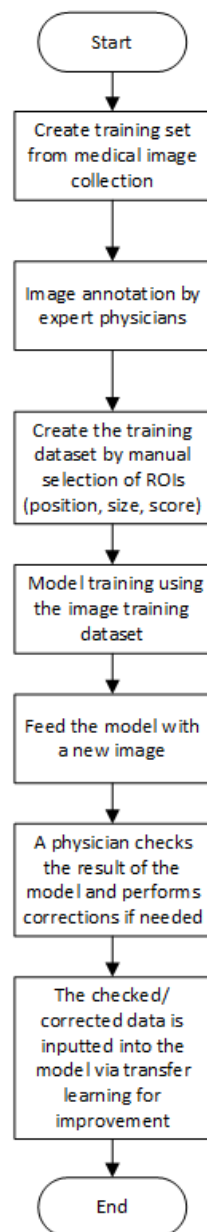


Figure 4. Block diagram for work of [10].

2.2. Datasets That Were Used

As mentioned above, the main modalities of medical images that were used were firstly the radiograph images (Figures 5 and 6) and secondly the ultrasound images (Figures 7 and 8). The authors used datasets that were either created from hospitals and medical centers or were benchmarked datasets such as the RA2 DREAM Challenge Dataset [29] and Medusa Database [30]. Regarding the custom datasets, the dimensions of the original medical images were varied. Furthermore, in some datasets, there were feet images included besides the hand images. According to the model, the original images were resized in order to fit the requirements. In works where there was ROIs detection (e.g., fingers, joints), there was a denoising process performed in order for the model to make the data as “clean” as possible. On the other hand, in works where there were no special ROIs detected, the data were loaded into the model in a raw form but sometimes modified (cropped image with only one hand [23], combination of four images into one showing both hands and feet [12]). It is important to notice that, as the patient data are protected under the scope of GDPR, all patients whose data were used to create datasets at

first agreed with the authorities of each institution for the usage of their data. However, even in that way, there were works with datasets that were too small; therefore, a data augmentation method was considered necessary so that efficient data could be created for the demanding training of the models. Moreover, the transfer learning method can solve the problem of limited datasets. Finally, it must be noticed that even with small datasets (e.g., <300 images), some authors achieved good results [16].

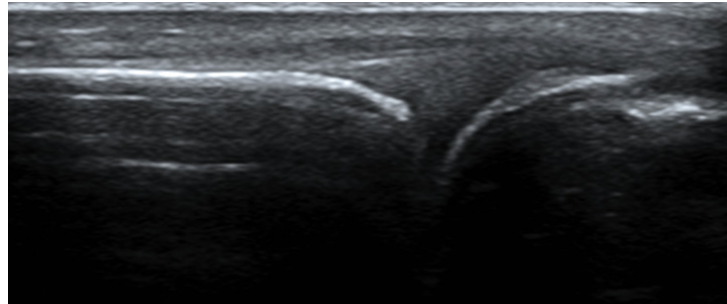


Figure 5. Example of a typical ultrasound image of a wrist used in the database of [15].

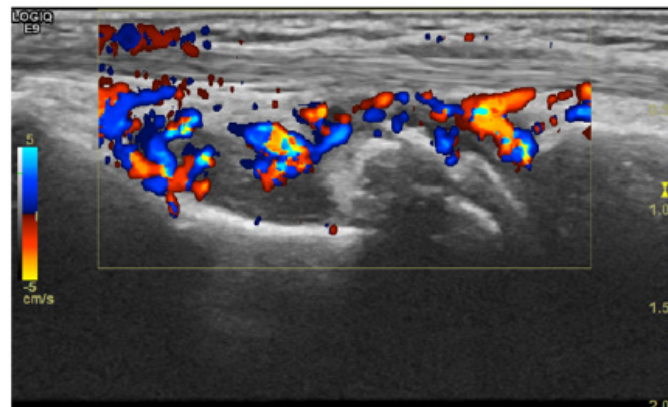


Figure 6. Example of an ultrasound color Doppler image of a wrist used in the database of [14].

2.3. The Physician Method

The diagnosis of RA is often complex, as it is based on the evaluation of many different parameters. These include patient history, physical examination, laboratory findings, X-ray-, ultrasound-findings and sometimes MRI findings.

2.3.1. Patient History

RA, as mentioned before, is a systemic disease, meaning that it can affect many organs and not only the joints. As far as joint involvement is concerned, typical for RA is symmetrical pain in the small joints, with swelling and morning stiffness, as well as improvement of symptoms during the day. Awakening in the night because of pain is also quite common. Inflammatory disorders cause pain that is usually noticed at rest and improves with movement. On the contrary, pain because of osteoarthritis or other degenerative disorders worsens with movement and is better at rest. When taking a medical history of a patient with RA, questions about possible extra-articular manifestations are of great importance. These questions include: trouble breathing (in case of interstitial lung involvement, pleural effusion, parenchymal pulmonary nodules), chest pain (in case of pericarditis, myocarditis, endocarditis, etc.), disturbance of vision as a result of eye inflammation (in case of scleritis/episcleritis/retinal vasculitis), palpable nodes under skin (rheumatoid nodules), other skin alterations (in the case of coexisting rheumatoid vasculitis = inflammation of the blood vessels because of RA), sicca symptoms meaning dryness of the eyes or mouth, often in combination with swelling of salivary glands (in case of

coexisting Sjögren syndrome), gastrointestinal symptoms such as abdominal pain, blood in the feces (in case of mesenteric vasculitis or intestinal infarction), pain when urination or change in the color of urine or any other kind of disturbance of renal function (in case of mesangial glomerulonephritis, amyloidosis, interstitial renal disease), weakness in the extremities, numbness or tingling, clumsiness and poor coordination of the hands (in case of peripheral neuropathy, mononeuritis multiplex, etc.), pain in the neck (in case of atlantoaxial instability) and tiredness/fatigue (in case of anemia on chronic disease because RA) [31]. In some cases, lymph node enlargement is also possible, mimicking Hodgkin’s disease [32].

2.3.2. Physical Examination

A well-known tool when coming to the regular evaluation of the course of RA patients in clinical praxis is the Disease Activity Score (DAS28). DAS28 has also been widely used in clinical trials when RA patients have been recruited in order to compare RA activity throughout the study. The DAS28 is a measure of RA disease activity that combines information from tender joints, swollen joints and inflammatory markers (CRP or erythrocyte sedimentation rate—ESR) [33]. The physician needs to press the below-marked 28 joints and mark the number of joints that are painful. CRP is the so-called C-reactive protein and ESR of the erythrocytes’ sedimentation rate. Both of them are measured in the blood and are high in the case of inflammation or infection. DAS28 can be calculated for both of them. Thus, there is the DAS28CRP and the DAS28ESR, respectively. Number 28 represents the 28 joints, which are typically being assessed in the clinical praxis by rheumatologists. These joints are presented on the mannequin below (Figure 7):

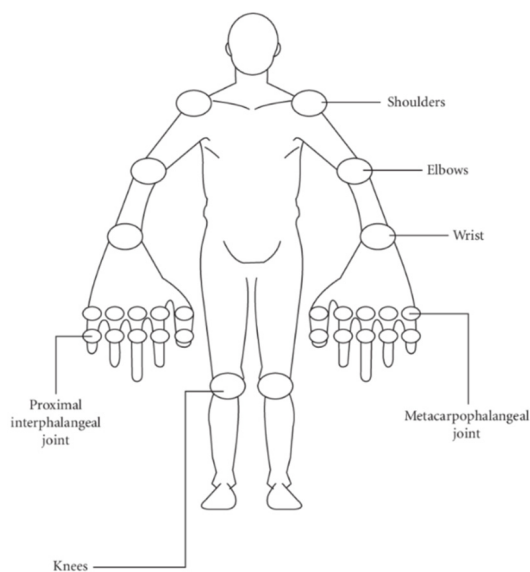


Figure 7. Self-administered index (SAI) modified from [34].

DAS28CRP or DAS28ESR can be easily calculated online [35].

In Table 1, RA activity in correlation with DAS28 score is presented.

Table 1. Discrimination of works based on deep learning model data input [36,37].

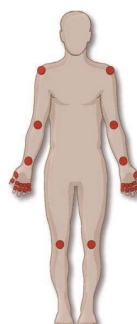
DAS28	RA Activity
DAS28 < 2.6	remission
2.6 ≤ DAS28 ≤ 3.2	low disease activity
3.2 < DAS28 ≤ 5.1	moderate disease activity
DAS28 > 5.1	high disease activity

Another widely used score for the clinical course of RA patients is CDAI (clinical disease activity index) (Figure 8). CDAI is calculated similarly. The main difference to DAS28 is inflammatory markers are not included; therefore, this score is more subjective. In CDAI, physicians are also called to quantify the disease activity on a scale from 0 to 10. CDAI is also based on tenderness and swelling of the same 28 peripheral joints.

Because of the possible extra-articular manifestations, it is always important that the physician examines the RA patient’s whole body and not just the joints. The skin, the eyes, heart and lungs, palpation for possible lymph nodes and abdomen also need to be thoroughly examined.

Clinical Disease Activity Index (CDAI)

Joint	Left		Right	
	Tender	Swollen	Tender	Swollen
Shoulder				
Elbow				
Wrist				
MCP 1				
MCP 2				
MCP 3				
MCP 4				
MCP 5				
PIP 1				
PIP 2				
PIP 3				
PIP 4				
PIP 5				
Knee				
Total	Tender:		Swollen:	



Patient Global Assessment of Disease Activity

Considering all the ways your arthritis affects you, rate how well you are doing on the following scale:

Very Well 0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10 Poor

Your Name _____ Date of Birth _____ Today's Date _____

Provider Global Assessment of Disease Activity

Very Well 0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10 Poor

How to Score the CDAI

Variable	Range	Value
Tender joint score	(0-28)	
Swollen joint score	(0-28)	
Patient global score	(0-10)	
Provider global score	(0-10)	
Add the above values to calculate the CDAI score	(0-76)	

CDAI Score Interpretation	
0.0 – 2.8	Remission
2.9 – 10.0	Low Activity
10.1 – 22.0	Moderate Activity
22.1 – 76.0	High Activity

Figure 8. Typical CDAI activity index template.

2.3.3. Laboratory Findings

In addition to inflammatory markers such as CRP and ESR, all possible RA patients must be screened for the presence of rheumatoid factor (RF) and anticitrullinated protein antibodies (ACPAs). When RF or ACPAs are positive, we can speak of seropositive RA. According to the literature, 70–80% of patients with RA are positive for autoantibodies, such as rheumatoid factors (RFs) and anti-citrullinated protein antibodies (ACPAs) [38]. Commonly, there are patients with rheumatoid arthritis with negative antibodies (RF, ACPA). In that case, the diagnosis of seronegative rheumatoid arthritis can be made.

2.3.4. X-ray Findings

RA has a predilection for MCP and PIP (proximal interphalangeal) joints, ulnar styloid and triquetrum (Figure 9). DIPs (distal interphalangeal joints) are spared.

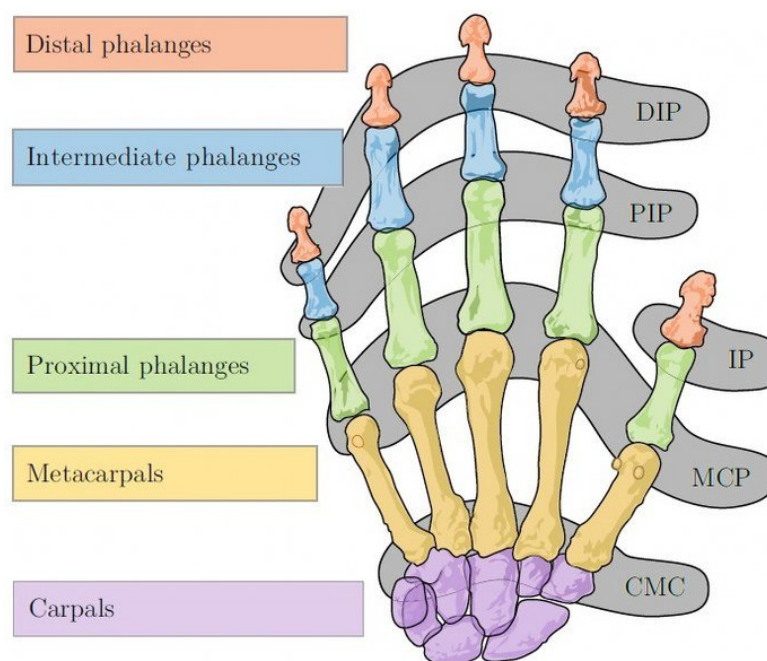


Figure 9. Bones and joints of the human hand, DIP—distal interphalangeal joint; PIP—proximal interphalangeal joint; IP—interphalangeal joint; MCP—metacarpophalangeal joint; CMC—carpo-metacarpal joint [39].

Typical X-ray findings in RA patients include marginal erosions, symmetrical joint space narrowing, subchondral cyst formation, subluxation causing ulnar deviation of MCP joints or boutonniere and swan neck deformities. Other typical X-ray findings in RA patients are hitchhiker’s thumb deformity, scapholunate dissociation, ulnar translocation, ankylosis (complete loss of the joints space) and scallop sign: erosion of the ulnar aspect of the distal radius, which may be predictive of extensor tendon rupture (Vaughan-Jackson syndrome).

Bone erosions, especially at the bare areas of joints (joint areas that are not covered with cartilage), are common for RA and are easily detectable with conventional X-rays [40].

There are many scoring systems that are used when evaluating an X-ray of a patient with RA. The most widely used nowadays is the modified Sharp/Van der Heijde [13] scoring system.

The original Sharp method assessed 27 joints in each hand and wrist, with each joint being given a separate score for joint space narrowing and erosions. The Sharp score focused on the hands and wrists (evaluation of 17 areas for erosions and 18 areas for joint space narrowing) and Van der Heijde added the feet in these evaluations, a modification that was also used by Sharp. Because of their similarities, these radiographic scoring systems are known as “modified Sharp methods.” In modified Sharp scoring systems, each joint is given a score for joint space narrowing and another score for erosions. Fifteen sites in each hand and wrist and six joints in each foot are examined for joint space narrowing on a scale of 0 to 4: 0 indicates no narrowing, 1 represents minimal narrowing, 2 indicates loss of 50% of the joint space, 3 indicates loss of 75% of the joint space, and 4 represents a complete loss of the joint space. The erosions are counted individually, usually at 16 sites in each hand and wrist and six sites in each foot. The erosion score per joint of the hands can range from 0 to 5. Erosions are scored 1 if they are discrete but clearly present and 2 or 3 if they are larger, depending on the surface area of the joint involved. A score of 4 is given if the erosion is large and extends over the imaginary middle of the bone. A score of 5 is given if a complete collapse of the joint is present or if the full surface of the joint is affected. In each joint, individual

erosions are summed up to a maximum of 5. The maximal erosion score for each hand is thus 80, considering the 16 areas for erosions per hand [41].

2.3.5. Ultrasound Findings

Musculoskeletal ultrasonography (MSUS) has established its role in the diagnosis of RA. High resolution gray scale (GSUS) and power Doppler (PDUS) assist the diagnostic performance of 2010 ACR (American College of Rheumatology)/EULAR (European League Against Rheumatism) classification criteria in early detection of RA.

Synovia is the inner layer of the articular capsule. It is a highly vascularized layer of serous connective tissue. It absorbs and secretes synovial fluid and is responsible for the mediation of nutrient exchange between blood and joint. In RA, a patient's ultrasound often reveals inflammation of the synovia, the so-called synovitis, with the typical accumulation of fluid in the synovia or synovial thickening. The presence of the power Doppler gives us information about the current activity of inflammation. Older inflammation sites most of the time do not have a power Doppler.

The Outcome Measures in Rheumatology (OMERACT) US Working Group formulated the definitions of pathological findings in ultrasound in RA patients and their quantification (Figure 10). The definition and grading of synovitis in RA are presented in Table 2 [17].

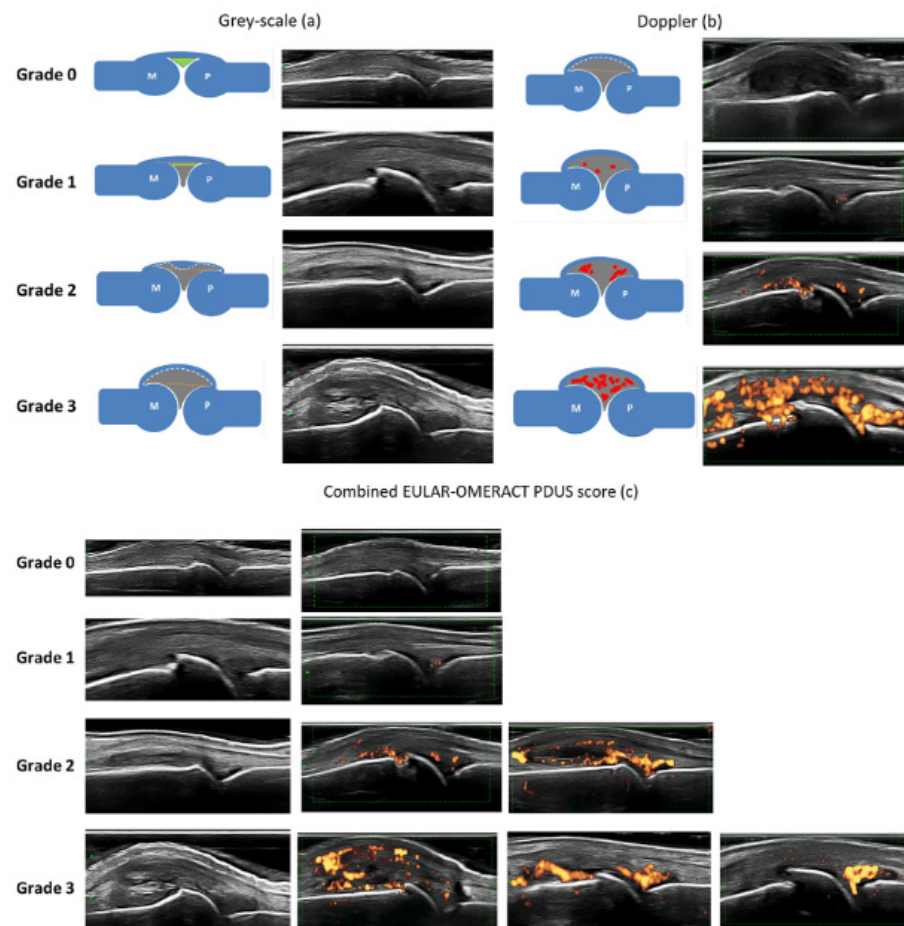


Figure 10. EULAR-OMERACT score examples [17].

Table 2. EULAR–OMERACT combined scoring system for grading synovitis in rheumatoid arthritis. (GS: gray scale, SH: synovial hypertrophy, PD: power Doppler).

Grade 0: Normal joint	No GS-detected SH and no PD signal within the synovium
Grade 1: Minimal synovitis	Grade 1 SH and \leq Grade 1 PD signal
Grade 2: Moderate synovitis	Grade 2 SH and \leq Grade 2 PW signal or Grade 1 SH and a Grade 2 PD signal
Grade 3: Severe synovitis	Grade 3 SH and \leq Grade 3 PD signal or Grade 1 or 2 SH and a Grade 3 PD signal

2.3.6. MRI Findings

MRI is a sensitive imaging modality that allows detailed assessment of inflammation as well as structural damage in RA. Compared to a physical examination, MRI is a more sensitive tool for the identification of tissue damages because of its direct visualization of synovitis, cartilage destruction, bone erosion, bone marrow edema, tenosynovitis, and surrounding soft tissue structures [40].

Synovitis (inflammation of the synovia) and tenosynovitis (inflammation of the tendons) at MCP (metacarpophalangeal joints), wrist and MTP (metatarsophalangeal joints) were independently associated with clinical swelling. MRI could detect inflammation in 54–64% of joints with no clinical swelling [42].

In order to quantify synovitis, bone marrow edema and bone erosions detected with MRI, an OMERACT RA MRI score (RAMPIS) has been developed and is often used for hands and wrists.

Overall, MRI is often used in the clinical praxis when there is diagnostic doubt in patients with clinically suspected arthralgia (CSA) because it can detect early findings of RA in the preclinical phase, thus offering the physician a window of opportunity of early treatment, which is the key for a good prognosis in RA patients.

2.3.7. The 2010 ACR—EULAR Classification Criteria for Rheumatoid Arthritis

Very useful for the classification of RA patients in clinical studies are the 2010 ACR—EULAR classification criteria for rheumatoid arthritis, which are presented below. A score of >6 is needed for a diagnosis of rheumatoid arthritis. However, it is important that they are classification and not diagnostic criteria. They are meant to be useful for the classification of RA patients in groups for clinical studies. The diagnosis of RA is based on a combination of clinical, radiographic, and serological findings, and the doctor is still the one who will make the diagnosis of RA [43].

- Joint involvement
 - o 0: 1 large joint
 - o 1: 2–10 large joints
 - o 2: 1–3 small joints (with or without the involvement of large joints)
 - o 3: 4–10 small joints (with or without the involvement of large joints)
 - o 5: >10 joints (at least 1 small joint)
- Serology
 - o 0: negative RF and negative anti-CCP
 - o 2: low-positive RF or low-positive anti-CCP
 - o 3: high-positive RF or high-positive anti-CCP
- Acute phase reactants
 - o 0: normal CRP and normal ESR
 - o 1: abnormal CRP and abnormal ESR
- Duration of symptoms
 - o 0: <6 weeks
 - o 1: >6 weeks

3. Results

We tried to classify the works that have been completed by examining several criteria. Thirteen out of fourteen studied works used only medical images in order to feed their deep learning models. Only one work [27] used additional patient data from EHR platforms that do not include medical images. Regarding the rest works, as it is shown in Table 3, the medical modalities that were used included radiograph medical images and ultrasound images. Most works were based on radiograph (X-ray) medical imaging. Another point of interest that differentiates the reviewed works is the existence or not of algorithms that focus on the detection and the localization of specific areas in the image. Hence, in some works, the authors fed the input image to the model without any segmentation actions, while others first extract specific ROIs, and these ROIs are finally driven in the deep learning model for classification. In all works, there has been a pre-processing stage where actions such as resizing, denoising and normalization took place in order to achieve the best outcome. Some authors developed their own deep learning models while others used as a base some of the already widely used models such as DenseNet, U-Net, LeNet-5, VGG16, ResNet-50 and Inception V3. The medical image datasets were mainly collected from medical centers and hospitals even if some authors used benchmarked datasets such as RA2 Dream Challenge Dataset [29] and Medusa Dataset [30]. The accuracies that have been noted by the models had a range from 50% to 98%. This deviation was associated with the type of images that were used, the pre-processing stage before the classification, the type of the models, and finally, the number of the output categories, meaning if the classification was binary (the model tried to predict if there is a RA disease or not), the accuracy of the models was high. On the other hand, if the models were asked to classify the image into one of the many possible categories that correspond to a scoring system (e.g., SvH [13] score proposes six categories), the accuracy was decreased.

The difficult field of the explainability and the trustworthiness of the deep learning models was not mainly counteracted by the researchers' community as only two works [14,25] dealt with the problem directly. It can be noted that, in addition to the above two works, four more works [10,11,21,25] faced the problem indirectly, by firstly detecting the joints in a medical image and then segmenting the image only using the joints areas as ROIs. The work of [24] should be added in the last group, as after the pre-processing stage in this work, only the areas of the bones of the fingers are loaded into the model.

The technology of transfer learning, which in some deep learning applications is used, was not used on a big scale in this field as only two works [24,28] tried this method with their models. In our opinion, this can be explained considering the two main possible problems of transfer learning: (1) the problem of negative transfer and (2) the problem of overfitting. The first problem is associated with the similarity between the two tasks, the initial task and the target task. If the new problem is not similar enough to the initial problem, the procedure of transfer learning may lead to worse results. Additionally, currently, there is no standard procedure to measure the similarity between two artificial intelligence (AI) problems. Regarding the second problem, it is not easy for the developer to choose which layers and weights should stay frozen during the new training in order to avoid overfitting. On the other hand, a problem can be solved with a sufficient but not huge number of input data, meaning that it is better to re-train the whole model with the limited but sufficient new dataset.

Table 3. Criteria analysis of conducted works based on image input.

Criteria	Methods	No. of Works	References
Type of medical image	Ultrasound Image	3	[14–16]
	Radiograph Image (X-ray)	10	[10–12,21–26,28]
Input analysis	Detection/Segmentation of ROIs	9	[10,11,16,21,22,24–26,28]
	Plain Input	4	[12,14,15,23]
Expainability Method	Direct (Attention maps)	2	[15,26]
	Indirect (Segmented ROIs)	4	[10,11,21,25]
	No Expainability	7	[12,14,16,22–24,28]
Augmentation	Yes	7	[12,21–25,28]
	No	6	[10,11,14–16,26]
Transfer learning	Yes	2	[24,28]
	No	11	[10–12,14–16,21–23,25,26]
Dataset	From Medical Center/Hospital	9	[10,14,15,21–25,28]
	Benchmarked	4	[11,12,16,26]

4. Discussion—Conclusions

At this point, we can summarize the results of our research. First of all, it must be noted that, in general, the usage of deep learning in RA disease diagnosis is limited. The reason for this situation can be considered the deviation that exists between the method that a physician follows and the approach of the RA problem that the deep models show. It should also be noticed that regarding all the works that have been reviewed, the part of the physician is significantly limited. In all these works, a physician is needed only for the training data labeling in the case that the data are not annotated and in some cases for checking the output of the model, i.e., comparing his scientific opinion with the result of the model. There is no prior knowledge with regard to the patient so there is no effort for collecting medical data that may help the diagnosis. Hence, in all these works, the diagnosis is based exclusively on the image modality that was used and its interpretation of each model. The attention is focused only on the image itself, and the attempts are limited to trying to analyze the patterns of the image in general or of specific ROIs in the best way.

On the other hand, the expert physician method is more holistic as it is not confined only to the image modalities. As mentioned above, the physician collects and uses various patient data in order to proceed with the diagnosis. The physician collects data from blood count results, medical history, medical drug use, performs real-time examinations by palpating the region of interest and finally proceeds with medical images, at first with the use of the radiograph image modality and secondly with ultrasound imaging. MRI may be the final option. Hence, it can be said that the physician uses a methodology with specific, well-oriented steps, which can be logically explained, in order to collect accumulative knowledge regarding the RA condition. Furthermore, the medical data accumulation is performed in such a way that it helps the final diagnosis in the sense that it is well-structured as it is not plain unorganized information. Therefore, this procedure of collecting organized medical data helps the physician to converge to a diagnosis. As the image modality analysis is included in the physician method too, we may notice that this expert-wise approach is a set that contains the deep learning approach as a subset.

On the other hand, all the works that have been reviewed use only parts from the medical data of the patient. The researchers mainly focus (93%) on medical images as data input for their models, and only one work (7%) [27] uses medical patient data in addition to medical images, as shown in Table 3. Therefore, it can be noticed that there was no work in which a combination of medical image data and remaining patient medical data were used together in a deep learning model. As it is shown in Table 3, it can be said that the explainability problem was not addressed, as only 15% of the works used a direct explainability method. The lack of an explanation of the model's result has a deep impact on the degree of trust that the model can achieve towards the patient or the doctor. Hence, we propose that more research should be carried out in the direction of combined data

input for deep learning models. Most of the data that a physician uses for a RA diagnosis should be taken into account for the creation of deep learning models that behave similarly to a physician, showing an expert-wise behavior. This could lead to the development of models with built-in trustworthiness that would be more acceptable from the end users. According to this perspective, we suggest that more explainability methods should be used to eliminate the black nature of the deep learning models and enhance the trust that end users show. Our future research will be guided under this scope.

Author Contributions: Conceptualization, G.P.A. and G.A.P.; methodology, G.P.A. and G.A.P.; investigation, G.P.A. and M.P.A.; resources, G.P.A.; data curation, G.P.A.; writing—original draft preparation, G.P.A. and M.P.A.; writing—review and editing, G.P.A. and G.A.P.; visualization, G.P.A.; supervision, G.A.P.; project administration, G.A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. England, B.R.; Ted, M.R. Epidemiology of, Risk Factors for, and Possible Causes of Rheumatoid Arthritis. UpToDate, 2020. Available online: <https://www.uptodate.com/contents/epidemiology-of-risk-factors-for-and-possible-causes-of-rheumatoid-arthritis#!> (accessed on 24 October 2021).
2. Cross, M.; Smith, E.; Hoy, D.; Carmona, L.; Wolfe, F.; Vos, T.; Williams, B.; Gabriel, S.; Lassere, M.; Johns, N.; et al. The global burden of rheumatoid arthritis: Estimates from the Global Burden of Disease 2010 study. *Ann. Rheum. Dis.* **2014**, *73*, 1316–1322. [[CrossRef](#)] [[PubMed](#)]
3. Crowson, C.S.; Matteson, E.L.; Myasoedova, E.; Michet, C.J.; Ernste, F.C.; Warrington, K.; Davis, J.; Hunder, G.G.; Thorneau, T.M.; Gabriel, S.E. The lifetime risk of adult-onset rheumatoid arthritis and other inflammatory autoimmune rheumatic diseases. *Arthritis Rheum.* **2010**, *63*, 633–639. [[CrossRef](#)] [[PubMed](#)]
4. Myasoedova, E.; Crowson, C.S.; Kremers, H.M.; Thorneau, T.M.; Gabriel, S.E. Is the incidence of rheumatoid arthritis rising? Results from Olmsted County, Minnesota, 1955–2007. *Arthritis Rheum.* **2010**, *62*, 1576–1582. [[CrossRef](#)] [[PubMed](#)]
5. Eriksson, J.K.; Neovius, M.; Ernestam, S.; Lindblad, S.; Simard, J.F.; Askling, J. Incidence of Rheumatoid Arthritis in Sweden: A Nationwide Population-Based Assessment of Incidence, Its Determinants, and Treatment Penetration: Assessment of RA Incidence in Sweden. *Arthritis Rheum.* **2013**, *65*, 870–878. [[CrossRef](#)]
6. Birnbaum, H.; Pike, C.; Kaufman, R.; Marynchenko, M.; Kidolezi, Y.; Cifaldi, M. Societal cost of rheumatoid arthritis patients in the US. *Curr. Med Res. Opin.* **2009**, *26*, 77–90. [[CrossRef](#)]
7. Zanis, L.; Nissen, M.J. Targeted Treatment in Spondyloarthritis. *Revue Medicale Suisse*. 11 March 2020. Available online: <https://www.ncbi.nlm.nih.gov/pubmed/32167249> (accessed on 24 October 2021).
8. Ker, J.; Wang, L.; Rao, J.; Lim, C.T. Deep Learning Applications in Medical Image Analysis. *IEEE Access* **2018**, *6*, 9375–9389. [[CrossRef](#)]
9. Tsakalidou, V.N.; Mitsou, P.; Papakostas, G.A. Computer vision in autoimmune diseases diagnosis—Current status and perspectives. In Proceedings of the 5th International Conference on Computational Vision and Bio Inspired Computing (ICCVBIC 2021), Coimbatore, Tamil Nadu, 25–26 November 2021; p. 16.
10. Hioki, Y.; Makino, K.; Koyama, K.; Haro, H.; Terada, H. Evaluation Method of Rheumatoid Arthritis by the X-ray Photograph using Deep Learning. In Proceedings of the 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech), Nara, Japan, 9–11 March 2021; pp. 444–447. [[CrossRef](#)]
11. Maziarz, K.; Krason, A.; Wojna, Z. Deep Learning for Rheumatoid Arthritis: Joint Detection and Damage Scoring in X-rays. *arXiv* **2021**, arXiv:2104.13915. Available online: <http://arxiv.org/abs/2104.13915> (accessed on 24 October 2021).
12. 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 (IRI), Las Vegas, NV, USA, 11–13 August 2020; pp. 399–402. [[CrossRef](#)]
13. Boini, S.; Guillemin, F. Radiographic scoring methods as outcome measures in rheumatoid arthritis: Properties and advantages. *Ann. Rheum. Dis.* **2001**, *60*, 817–827.
14. Andersen, J.K.H.; Pedersen, J.S.; Laursen, M.S.; Holtz, K.; Grauslund, J.; Savarimuthu, T.R.; Just, S.A. Neural networks for automatic scoring of arthritis disease activity on ultrasound images. *RMD Open* **2019**, *5*, e000891. [[CrossRef](#)]
15. Dong, F.; Liu, Y.; Cui, C.; Shi, S.; Zeng, J.; Zhang, Y. A Deep Learning Classification of Metacarpophalangeal Synovial Proliferation in Rheumatoid Arthritis by Ultrasound Images. *SSRN J.* **2020**. [[CrossRef](#)]

16. Hemalatha, R.J.; Vijaybaskar, V.; Thamizhvani, T.R. Automatic localization of anatomical regions in medical ultrasound images of rheumatoid arthritis using deep learning. *Proc. Inst. Mech. Eng. Part H J. Eng. Med.* **2019**, *233*, 657–667. [[CrossRef](#)]
17. D’Agostino, M.-A.; Terslev, L.; Aegerter, P.; Backhaus, M.R.; Balint, P.; A Bruyn, G.; Filippucci, E.; Grassi, W.; Iagnocco, A.; Jousse-Joulin, S.; et al. Scoring ultrasound synovitis in rheumatoid arthritis: A EULAR-OMERACT ultrasound taskforce—Part 1: Definition and development of a standardised, consensus-based scoring system. *RMD Open* **2017**, *3*, e000428. [[CrossRef](#)]
18. Singh, A.; Sengupta, S.; Lakshminarayanan, V. Explainable Deep Learning Models in Medical Image Analysis. *J. Imaging* **2020**, *6*, 52. [[CrossRef](#)]
19. Maier, A.; Syben, C.; Lasser, T.; Riess, C. A gentle introduction to deep learning in medical image processing. *Z. Med. Phys.* **2019**, *29*, 86–101. [[CrossRef](#)]
20. London, A.J. Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Cent. Rep.* **2019**, *49*, 15–21. [[CrossRef](#)]
21. Hirano, T.; Nishide, M.; Nonaka, N.; Seita, J.; Ebina, K.; Sakurada, K.; Kumanogoh, A. Development and validation of a deep-learning model for scoring of radiographic finger joint destruction in rheumatoid arthritis. *Rheumatol. Adv. Pr.* **2019**, *3*, rkz047. [[CrossRef](#)]
22. Huang, Y.-J.; Shun, M.; Zheng, K.; Lu, L.; Lu, Y.; Lin, C.; Kuo, C.-F. Radiographic Bone Texture Analysis Using Deep Learning Models for Early Rheumatoid Arthritis Diagnosis. 2020. Available online: <https://assets.researchsquare.com/files/rs-76193/v1/415a2e2e-13c1-4575-ac79-c4e743cf7307.pdf?c=1631855828> (accessed on 24 October 2021).
23. Üreten, K.; Erbay, H.; Maraş, H.H. Detection of rheumatoid arthritis from hand radiographs using a convolutional neural network. *Clin. Rheumatol.* **2019**, *39*, 969–974. [[CrossRef](#)]
24. Murakami, S.; Hatano, K.; Tan, J.; Kim, H.; Aoki, T. Automatic identification of bone erosions in rheumatoid arthritis from hand radiographs based on deep convolutional neural network. *Multimed. Tools Appl.* **2018**, *77*, 10921–10937. [[CrossRef](#)]
25. Lee, S.; Choi, M.; Choi, H.-S.; Park, M.S.; Yoon, S. FingerNet: Deep learning-based robust finger joint detection from radiographs. In Proceedings of the 2015 IEEE Biomedical Circuits and Systems Conference (BioCAS), Atlanta, GA, USA, 22–24 October 2015; pp. 1–4. [[CrossRef](#)]
26. Chaturvedi, N. DeepRA: Predicting Joint Damage From Radiographs Using CNN with Attention. *arXiv* **2021**, arXiv:2102.06982. Available online: <http://arxiv.org/abs/2102.06982> (accessed on 24 October 2021).
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](#)]
28. Rohrbach, J.; Reinhard, T.; Sick, B.; Dürr, O. Bone erosion scoring for rheumatoid arthritis with deep convolutional neural networks. *Comput. Electr. Eng.* **2019**, *78*, 472–481. [[CrossRef](#)]
29. Dream Challenge: Automated Scoring of Radiographic Joint Damage. 2020. Available online: <https://www.synapse.org/#/Synapse:syn20545111/wiki/594083> (accessed on 29 October 2021).
30. Mielnik, P.; Fojcik, M.; Segen, J.; Kulbacki, M. A novel method of synovitis stratification in ultrasound using machine learning algorithms: Results from clinical validation of the Medusa Project. *Ultrasound Med. Biol.* **2018**, *44*, 489–494. [[CrossRef](#)]
31. Marcucci, E.; Bartoloni, E.; Alunno, A.; Valentini, V.; Valentini, E.; La Paglia, G.M.C.; Bonifacio, A.F.; Gerli, R. Extra-articular rheumatoid arthritis. *Reumatismo* **2018**, *70*, 212–224. [[CrossRef](#)]
32. Das, S.; Padhan, P. An Overview of the Extraarticular Involvement in Rheumatoid Arthritis and its Management. *J. Pharmacol. Pharmacother.* **2017**, *8*, 81–86.
33. Van Riel, P.L.C.M.; Renskers, L. The Disease Activity Score (DAS) and the Disease Activity Score using 28 joint counts (DAS28) in the management of rheumatoid arthritis. *Clin. Exp. Rheumatol.* **2016**, *34*, S40–S44.
34. Amaya-Amaya, J.; Botello-Corzo, D.; Calixto, O.-J.; Calderón-Rojas, R.; Domínguez, A.-M.; Cruz-Tapias, P.; Montoya-Ortiz, G.; Mantilla, R.-D.; Anaya, J.-M.; Rojas-Villarraga, A. Usefulness of Patients-Reported Outcomes in Rheumatoid Arthritis Focus Group. *Arthritis* **2012**, *2012*, 1–13. [[CrossRef](#)]
35. DAS 28—Disease Activity Score Calculator for Rheumatoid Arthritis. Available online: <http://www.4s-dawn.com/DAS28/> (accessed on 30 October 2021).
36. Fleischmann, R.M.; Van Der Heijde, D.; Gardiner, P.; Szumski, A.; Marshall, L.; Bananis, E. DAS28-CRP and DAS28-ESR cut-offs for high disease activity in rheumatoid arthritis are not interchangeable. *RMD Open* **2017**, *3*, e000382. [[CrossRef](#)]
37. Fransen, F. Remission in rheumatoid arthritis: Agreement of the disease activity score (DAS28) with the ARA preliminary remission criteria. *Rheumatology* **2004**, *43*, 1252–1255. [[CrossRef](#)]
38. Smolen, J.S.; Aletaha, D.; Barton, A.; Burmester, G.R.; Emery, P.; Firestein, G.S.; Kavanaugh, A.; McInnes, I.B.; Solomon, D.H.; Strand, V.; et al. Rheumatoid Arthritis. Nature Reviews. Disease Primers. Available online: <https://pubmed.ncbi.nlm.nih.gov/29417936/> (accessed on 30 October 2021).
39. Tavakoli, M.; Batista, R.; Sgrigna, L. The UC SoftHand: Light Weight Adaptive Bionic Hand with a Compact Twisted String Actuation System. *Actuators* **2015**, *5*, 1. [[CrossRef](#)]

40. Zabotti, A.; Finzel, S.; Baraliakos, X.; Aouad, K.; Ziade, N.; Iagnocco, A. Review Imaging in the preclinical phases of rheumatoid arthritis. *Clin. Exp. Rheumatol.* **2020**, *38*, 536–542.
41. Ory, P.A. Interpreting radiographic data in rheumatoid arthritis. *Ann. Rheum. Dis.* **2003**, *62*, 597–604. [[CrossRef](#)] [[PubMed](#)]
42. Mathew, A.J.; Danda, D.; Conaghan, P.G. MRI and ultrasound in rheumatoid arthritis. *Curr. Opin. Rheumatol.* **2016**, *28*, 323–329. [[CrossRef](#)] [[PubMed](#)]
43. Kay, J.; Upchurch, K.S. ACR/EULAR 2010 rheumatoid arthritis classification criteria. *Rheumatology* **2012**, *51*, vi5–vi9. [[CrossRef](#)] [[PubMed](#)]