

Early Diagnosis of Ankylosing Spondylitis using MRI and Transfer Learning

Mohit Kumar¹, Piyush Patil², Yash Singh³ and Tanmayee Patki⁴

¹⁻⁴MIT-SOC, Pune, India

Email: mohit.kumar@mituniversity.edu.in

Abstract— Ankylosing spondylitis (AS) is a chronic inflammatory arthritis that mainly affects the spine and sacroiliac (SI) joints. AS causes severe pain and stiffness, which can greatly limit a person's ability to perform daily activities. Early diagnosis of AS is important for effective treatment and to prevent or delay irreversible damage. However, due to the rarity of the disease and the lack of available data, developing accurate diagnostic tools remains a challenge. AS patients are frequently misdiagnosed due to the lack of awareness about AS and the overlapping symptoms with other common conditions. In this research, we propose a method for the early diagnosis of ankylosing spondylitis using MRI scans of the SI joint. Given the limited dataset available for AS, we use transfer learning techniques to leverage knowledge from a large dataset of sacroiliitis MRI scans, which may not necessarily be caused by AS. Specifically, we fine-tune a pre-trained model, originally trained to classify a diverse set of arthritis cases as sacroiliitis or not, to better recognize AS-specific patterns in the SI joint MRI images. Our approach has achieved an accuracy of 96%. Being able to detect and differentiate AS from other similar conditions, among which it is often misdiagnosed, represents a significant advancement in the diagnosis of AS.

Index Terms—Ankylosing Spondylitis, Transfer Learning, Medical Image Classification, Magnetic Resonance Imaging.

I. INTRODUCTION

Ankylosing spondylitis (AS) is a chronic condition that causes inflammation primarily in the spine and sacroiliac joints (SI). It's known for its distinctive: inflammatory lower back pain that often worsens with rest and improves with activity. Despite this, AS is frequently misdiagnosed, leading to significant delays in getting the right treatment in time. The cause for this disease is still unknown although genetics, particularly the HLA-B27 gene, play a role, environmental factors also contribute. One of the biggest hurdles in diagnosing AS is the time it takes to confirm diagnosis, which can average between 8 to 11 years from when symptoms first appear. Several reasons contribute to this delay, including the gradual onset of symptoms, which can be mistaken for typical back pain, especially in the early stages. Many healthcare providers may not be familiar with AS to recognize its symptoms, particularly in primary care settings. Most people are not familiar with rheumatologists and do not reach the right specialist for their disease. Traditionally, X-rays have been the method for diagnosing AS, focusing on structural changes in the sacroiliac joints and spine. However, X-rays often miss the early signs of inflammation in the SI joint. MRI scans can pick up on subtle inflammatory changes, even before structural damage occurs.

This early detection can help treatment decisions and potentially prevent or delay irreversible damage. The power of artificial intelligence (AI) can be harnessed for AS diagnosis. AI classifiers, trained on vast amounts of data, can identify patterns and abnormalities in MRI scans that might be missed by the human eye. By combining these AI insights with patient symptoms and medical history, healthcare providers can make more informed decisions, potentially speeding up diagnosis and treatment. AI will enhance but not replace, the skills of experienced healthcare professionals. Diagnosing AS early is crucial for preventing long-term complications, delaying progression of disease, and improving quality of life. Advancements in AI technology, hold significant potential for speeding up diagnosis and getting patients the treatment, they need. By raising awareness about the disease and using innovative tools, we can reduce diagnosis time and improve quality of life for those who are living with AS.

A. Literature review

In recent literature, various studies have explored the diagnosis of diseases using MR images [7–11]. Specifically, research focused on ankylosing spondylitis (AS) includes Han et al. [12], who developed an automated algorithm using MRI for quantifying and grading AS and hip arthritis. This deep learning-based system identifies inflammatory regions and evaluates severity levels in MRIs. The algorithm reported accuracies of 85.7%, distinguishing between grades of bone marrow edema. Navarani et al. [13] explored the effectiveness of machine learning in assessing cardiovascular risks in AS patients. Li et al. [12] introduced an AI-based system for the diagnosis and treatment of AS, employing an ensemble deep learning model. Zunti et al. [14] used medical imaging, specifically CT scans, with machine learning to detect erosion, an early indicator of AS. Lin et al. [15] applied a deep learning algorithm using STIR sequences in MRIs to detect active inflammatory sacroiliitis, involving 326 patients diagnosed with axial spondylarthritis and 63 with non-specific back pain. The algorithm matched a radiologist's sensitivity and specificity and surpassed the performance of a rheumatologist, confirming its potential for diagnosing SA and assessing disease activity. Bressem et al. [16] employed an artificial neural network (ANN) to identify sacroiliitis in axial spondylarthritis using conventional radiographs of the sacroiliac joints. Involving 2011 radiographs, the ANN showed outstanding performance, with AUCs of 97.00% and 94.00% for validation and test datasets, respectively. Shenkman et al. [17] used a supervised machine learning algorithm for detecting and grading sacroiliitis in CT scans. Bressem et al. [18] and Nevsun et al. [19] reported significant advancements using deep learning for diagnosing axial spondylarthritis and AS, respectively, with Nevsun's ASNET achieving a 99.8% accuracy on their dataset.

II. THE PROPOSED METHOD

In this study we propose a transfer learning approach for the detection of AS. As the disease is rare and getting diagnosed at an early stage is rarer, we found out that creating a large dataset is extremely difficult. We created a large dataset of sacroiliitis caused by various reasons for the initial training of the model. Then finetuned the same model on a dataset containing MRI of SI joints of AS patients. This improves accuracy for small datasets, this can be very helpful in other rare diseases with less research and awareness.

A. The dataset

Two datasets were utilized in this study:

- General Sacroiliitis Dataset: Comprising 1,000 MRI scans of the SI joint from various inflammatory causes, not limited to AS.
- Specific AS Dataset: Consisting of 300 MRI scans from confirmed AS patients.



Figure 1. MRI of SI joint of AS patient (on the left) and of healthy patient (on the right)

Both datasets underwent preprocessing, including greyscale conversion and resizing to a uniform dimension of 1000x1000 pixels to facilitate efficient processing by the CNN.

B. The proposed method

We employed a two-phase training approach using a CNN architecture:

Training Phase 1 - Generalization

The model is trained using a learning rate of 1e-4 and batch size of 32 on the general sacroiliitis dataset. This training aims to create a model which can handle variety of cases of SI joint MRI and distinguish between healthy and unhealthy patients. With the completion of this training our model can identify the region of interest (ROI) in the SI joint, and also is able to identify corroded articular surfaces. The model is also able to distinguish between healthy SI joints and inflamed SI joints using MRI.

Training Phase 2 – Fine tuning

This is the finetuning phase, here the pretrained model will be trained again, but this time on the AS dataset with a reduced learning rate of 1e-6 and batch size of 16. This phase aims to train the model to differentiate AS from other diseases as well as healthy MRI. The previous phase helps the model to train efficiently and avoid overfitting before attaining good validation accuracy.

Layer (type)	Output Shape	Param #
conv2d_9 (Conv2D)	(None, 998, 998, 16)	160
max_pooling2d_4 (MaxPooling2D)	(None, 499, 499, 16)	0
conv2d_10 (Conv2D)	(None, 497, 497, 32)	4,640
conv2d_11 (Conv2D)	(None, 495, 495, 64)	18,496
max_pooling2d_5 (MaxPooling2D)	(None, 247, 247, 64)	0
max_pooling2d_6 (MaxPooling2D)	(None, 123, 123, 64)	0
conv2d_12 (Conv2D)	(None, 121, 121, 128)	73,856
conv2d_13 (Conv2D)	(None, 119, 119, 128)	147,584
max_pooling2d_7 (MaxPooling2D)	(None, 59, 59, 128)	0
max_pooling2d_8 (MaxPooling2D)	(None, 29, 29, 128)	0
conv2d_14 (Conv2D)	(None, 27, 27, 256)	295,168
conv2d_15 (Conv2D)	(None, 25, 25, 256)	590,080
max_pooling2d_9 (MaxPooling2D)	(None, 12, 12, 256)	0
max_pooling2d_10 (MaxPooling2D)	(None, 6, 6, 256)	0
flatten_1 (Flatten)	(None, 9216)	0
dense_2 (Dense)	(None, 32)	294,944
dense_3 (Dense)	(None, 2)	66

Total params: 1,424,994 (5.44 MB)
Trainable params: 1,424,994 (5.44 MB)
Non-trainable params: 0 (0.00 B)

Figure 2. Model Summary

This model is designed to scale down images of 1000x1000 pixels while extracting useful characteristics. Max Pooling layers are used to reduce the scale of the image after few Convolutional operations. This method is repeated until the images are scaled down to 256 6x6 pixel images. Then the images are flattened and sent to the Dense layers. The Dense layers process the inputs to output 2 floating point numbers. These numbers are interpreted as confidence of the model for the two classes.

III. EVALUATION

A. Experimental Design

To validate the effectiveness of our proposed transfer learning approach, we meticulously designed an experimental framework that compares the performance of our model against conventional training methodologies. Specifically, we focused on two baseline models:

- Direct CNN Training on AS Dataset: This model was trained solely on the AS-specific dataset without leveraging any prior knowledge or pre-training. The purpose of this baseline is to simulate a scenario where a model is developed specifically for AS using a limited dataset, representing a common challenge in medical imaging tasks where specialized data is scarce.

- **ResNet Pre-Trained Model:** For this, we used a ResNet architecture, which is widely recognized for its deep residual learning framework. We used the same dense layers as in the proposed models after the convolutional layers. The ResNet model was also trained directly on the AS dataset. Given ResNet's capabilities in handling complex image recognition tasks, this comparison was intended to compare the performance with our proposed model.
- **DenseNet Pre-Trained Model:** For this, we used a DenseNet architecture, which is widely known for its picture classification capabilities. We used the same dense layers as in the proposed models after the convolutional layers. The DenseNet model was also trained directly on the AS dataset. This comparison was intended to compare the performance with our proposed model.

B. Training Details

Both baseline models and the proposed transfer learning model followed the same data preprocessing steps, which included resizing images to 1000x1000 pixels and converting them to greyscale. This standardization ensured that any observed performance differences could be attributed to the model architecture and training strategy rather than external factors.

- **Direct CNN Model:** Trained from scratch with a learning rate of 1e-4, using the proposed CNN architecture.
- **ResNet Model:** Implemented using a pre-trained ResNet-50 architecture with the final layers adapted to our specific task. The model was fine-tuned with a learning rate of 1e-4 to adjust the deep layers to the specific dataset of the AS MRIs.
- **DenseNet Model:** Implemented using a pre-trained DenseNet model with the final layers adapted to our specific task. The model was fine-tuned with a learning rate of 1e-4 to adjust the dense layers to the specific dataset of the AS MRIs.
- **Proposed Transfer Learning Model:** Initiated training on the general sacroiliitis dataset with a learning rate of 1e-4, followed by fine-tuning on the AS-specific dataset at a reduced learning rate of 1e-6.

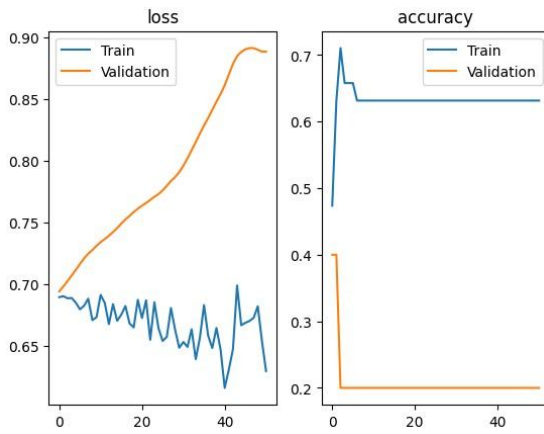


Figure 4. Training performance of CNN Model

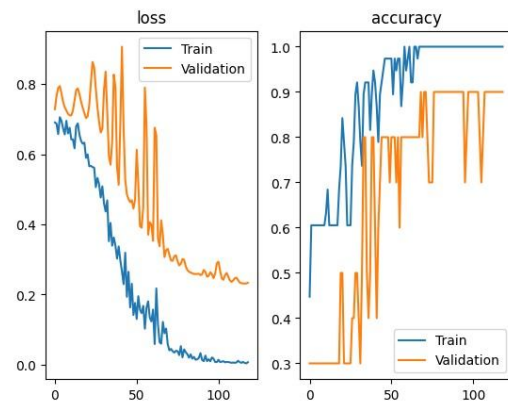


Figure 5. Training performance of ResNet Model

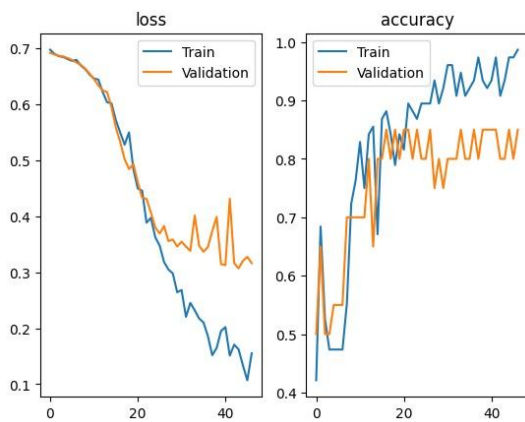


Figure 6. Training performance of DenseNet Model

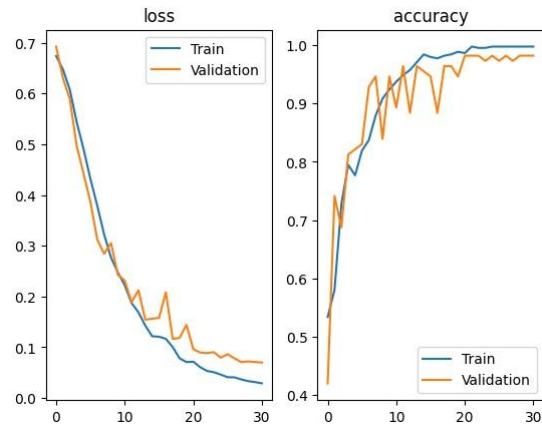


Figure 7. Training performance of Proposed Model

C. Evaluation Metrics

The primary metrics used to evaluate and compare the performance of each model were accuracy, sensitivity (recall), and specificity. These metrics are particularly pertinent in medical diagnostic applications where both the identification of true positives and the accurate exclusion of negatives are crucial. The definitions are as follows:

- Accuracy: The proportion of true results (both true positives and true negatives) in the total population.
- Sensitivity: The ability of the model to correctly identify patients with AS (true positive rate).
- Specificity: The ability of the model to correctly identify patients without AS (true negative rate).

D. Results

The results from the experiments were as follows:

- Direct CNN Model achieved an accuracy of 51%, sensitivity of 59%, and specificity of 53%.

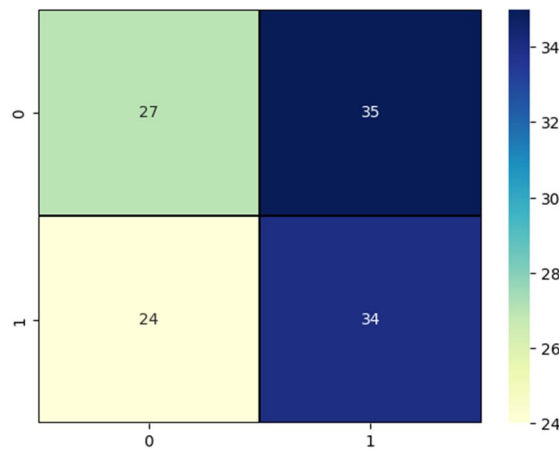


Figure 8. Confusion Matrix of CNN

	precision	recall	f1-score	support
0	0.53	0.44	0.48	62
1	0.49	0.59	0.54	58
accuracy			0.51	120
macro avg	0.51	0.51	0.51	120
weighted avg	0.51	0.51	0.51	120

Figure 9. Classification Report of CNN Model

- ResNet Model showed improved performance with an accuracy of 88%, sensitivity of 90%, and specificity of 87%.

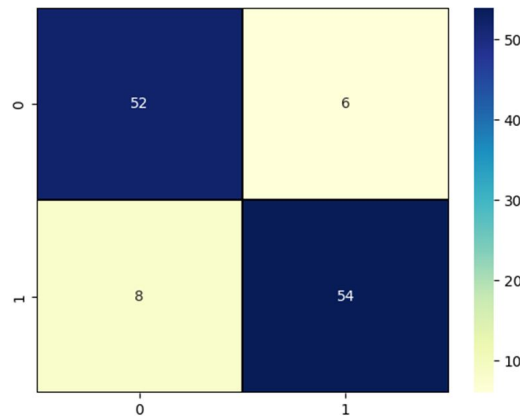


Figure 10. Confusion Matrix of ResNet Model

	precision	recall	f1-score	support
0	0.87	0.90	0.88	58
1	0.90	0.87	0.89	62
accuracy			0.88	120
macro avg	0.88	0.88	0.88	120
weighted avg	0.88	0.88	0.88	120

Figure 11. Classification Report of ResNet Model

- DenseNet Model showed good performance with an accuracy of 82%, sensitivity of 80%, and specificity of 84%.

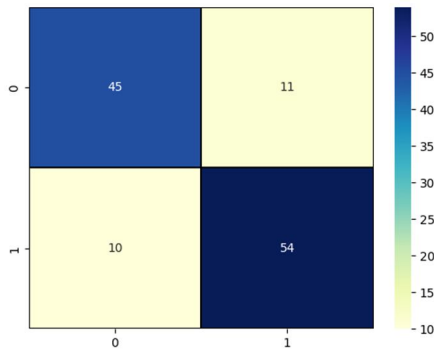


Figure 12. Confusion Matrix of DenseNet Model

	precision	recall	f1-score	support
0	0.82	0.80	0.81	56
1	0.83	0.84	0.84	64
accuracy			0.82	120
macro avg	0.82	0.82	0.82	120
weighted avg	0.82	0.82	0.82	120

Figure 13. Classification Report of DenseNet Model

The proposed Transfer Learning Model outperformed both baselines, attaining an accuracy of 96%, sensitivity of 98%, and specificity of 94%.

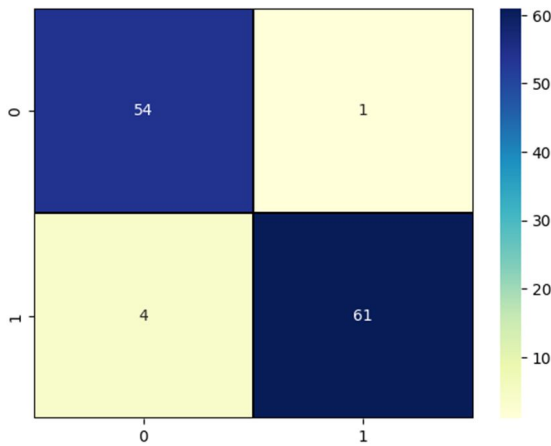


Figure 14. Confusion Matrix of Proposed Model

	precision	recall	f1-score	support
0	0.93	0.98	0.96	55
1	0.98	0.94	0.96	65
accuracy			0.96	120
macro avg	0.96	0.96	0.96	120
weighted avg	0.96	0.96	0.96	120

Figure 15. Classification Report of Proposed Model

These results show the advantage of using a transfer learning approach in scenarios where direct training on a specialized, limited dataset might cause overfitting or undertraining.

The initial training on a more diverse, general dataset allowed our model to develop understanding of the SI joint and identify the region of interest, which when fine-tuned to the AS-specific dataset, improved its ability to accurately classify AS conditions. Also, we showed that general pre-trained models trained on unrelated datasets are not best suited for specific scenarios.

This not only highlights the utility of transfer learning in medical imaging but also demonstrates its potential in improving diagnosis accuracy for rare diseases like AS.

IV. DISCUSSION

The findings of this study show the potential of transfer learning to enhance the diagnostic capabilities of machine learning models in medical image classification, especially for conditions such as ankylosing spondylitis (AS), where data availability is limited, and early detection is rare.

The training approach used in this study shows several benefits and implications for both AS diagnosis and the broader context of rare diseases.

A. Benefits of Transfer Learning in Medical Image classification

Transfer learning offers several advantages in medical image classification tasks:

- **Optimized Model Performance:** Transfer learning allows for the utilization of important features learned from related tasks, leading to improved model performance.

- **Efficient Use of Limited Data:** In the case of AS and other rare diseases, where obtaining large, annotated datasets is challenging, transfer learning enables the use of available data by adapting models trained on larger datasets to specific conditions.
- **Reduced Training Time and Resources:** Transfer learning requires less computational resources and training time compared to training models from scratch, making it a more practical approach, especially in medical research where resources are limited.

B. Challenge of Scarce Data in AS Diagnosis

Accurate diagnosis of AS is a challenge due to the scarcity of available data and the complexity of early symptoms:

- **Limited Dataset Availability:** The rarity of AS and the delay in diagnosis result in limited annotated datasets for training machine learning models. This scarcity hampers the development of accurate diagnostic tools, hindering efforts to improve patient outcomes.
- **Complexity of Disease Progression:** AS is a complex disease with varied symptoms and progression patterns, making it challenging to accurately identify specific biomarkers for early detection. This complexity further increases the difficulty of training reliable machine learning models.

C. Implications for Other Rare Diseases

The findings of this study have broader implications beyond AS, particularly for other rare diseases facing similar scarcity in data:

- **Generalizability:** The transfer learning approach employed in this study can be applied to other rare diseases where data is limited. By using pre-trained models trained on other related disease datasets and adapting them to specific conditions, this method can improve accuracy and reduce training time of models significantly.
- **Research and Awareness:** By highlighting the importance of innovative approaches like transfer learning in addressing diagnostic challenges, this study sheds light on the need for increased research and awareness for rare diseases. Collaborative efforts among researchers and healthcare institutions are essential for advancing diagnostic techniques and improving patient quality of life.

Transfer learning represents a powerful tool for improving the early diagnosis of rare diseases such as AS, where data scarcity is the biggest challenge for training accurate models. By leveraging existing knowledge from similar diseases and then adapting it to a specific disease.

V. CONCLUSIONS

In conclusion, this study provides a promising approach for transfer learning in the medical imaging field to improve the early diagnosis of rare diseases such as AS and other rare conditions where data availability is limited due to the difficulty in early diagnosis.

Using a transfer learning approach, we have demonstrated that diagnostic accuracy and efficiency can be improved, especially for AS, where early diagnosis is important for timely intervention. To overcome the challenge of data shortage in AS, our proposed approach uses a larger general dataset of SI joint which is the region of interest. Our model develops a deep understanding of related diseases through a dual phase training approach, which allows our model to easily understand the patterns present in the limited AS dataset.

We proved that models trained on related diseases outperform popular pretrained models trained on general image classification datasets. This reduces computational power as the proposed model can be a lot smaller than other pretrained models for image classification. The benefits of transfer learning for rare diseases go beyond AS and offer a promising approach to improve diagnostics for other rare diseases. Combined efforts of researchers can further improve and create innovative ideas for diagnosis of AS and other rare diseases.

ACKNOWLEDGMENT

The authors wish to thank A, B, C. This work was supported in part by a grant from XYZ.

REFERENCES

- [1] Braun, J.; Sieper, J. Ankylosing spondylitis. *Lancet* 2007, 369, 1379–1390. [CrossRef] [PubMed]
- [2] Xi, Y.; Jiang, T.; Chaurasiya, B.; Zhou, Y.; Yu, J.; Wen, J.; Shen, Y.; Ye, X.; Webster, T.J. Advances in nanomedicine for the treatment of ankylosing spondylitis. *Int. J. Nanomed.* 2019, 14, 8521–8542. [CrossRef] [PubMed]

- [3] Salvadorini, G.; Bandinelli, F.; Sedie, A.D.; Riente, L.; Candelieri, A.; Generini, S.; Possemato, N.; Bombardieri, S.; Matucci Cerinic, M. Ankylosing spondylitis: How diagnostic and therapeutic delay have changed over the last six decades. *Clin. Exp. Rheumatol.-Incl Suppl.* 2012, 30, 561.
- [4] Ritchlin, C.; Adamopoulos, I.E. Adamopoulos Axial spondyloarthritis: New advances in diagnosis and management. *BMJ* 2021, 372, m4447. [CrossRef]
- [5] Schueller-Weidekamm, C. Entzündliche Wirbelsäulenerkrankungen: Spondylarthritis. *Der Radiol.* 2015, 4, 337–348. [CrossRef] [PubMed]
- [6] Ou, J.; Xiao, M.; Huang, Y.; Tu, L.; Chen, Z.; Cao, S.; Wei, Q.; Gu, J. Serum metabolomics signatures associated with ankylosing spondylitis and TNF inhibitor therapy. *Front. Immunol.* 2021, 12, 630791. [CrossRef] [PubMed]
- [7] Kaplan, E.; Baygin, M.; Barua, P.D.; Dogan, S.; Tuncer, T.; Altunisik, E.; Palmer, E.E.; Acharya, U.R. ExHiF: Alzheimer's disease detection using exemplar histogram-based features with CT and MR images. *Med. Eng. Phys.* 2023, 115, 103971. [CrossRef] [PubMed]
- [8] Poyraz, A.K.; Dogan, S.; Akbal, E.; Tuncer, T. Automated brain disease classification using exemplar deep features. *Biomed. Signal Process. Control* 2022, 73, 103448. [CrossRef]
- [9] Kaplan, E.; Altunisik, E.; Firat, Y.E.; Barua, P.D.; Dogan, S.; Baygin, M.; Demir, F.B.; Tuncer, T.; Elizabeth Emma, P.; Tan, R.-S.; et al. Novel nested patch-based feature extraction model for automated Parkinson's Disease symptom classification using MRI images. *Comput. Methods Programs Biomed.* 2022, 224, 107030. [CrossRef]
- [10] Kaplan, E.; Dogan, S.; Tuncer, T.; Baygin, M.; Altunisik, E. Altunisik Feed-forward LPQNet based automatic alzheimer's disease detection model. *Comput. Biol. Med.* 2021, 137, 104828. [CrossRef]
- [11] Macin, G.; Tasci, B.; Tasci, I.; Faust, O.; Barua, P.D.; Dogan, S.; Tuncer, T.; Tan, R.-S.; Acharya, U.R. An accurate multiple sclerosis detection model based on exemplar multiple parameters local phase quantization: ExMPLPQ. *Appl. Sci.* 2022, 12, 4920. [CrossRef]
- [12] Han, Q.; Lu, Y.; Han, J.; Luo, A.; Huang, L.; Ding, J.; Zhang, K.; Zheng, Z.; Jia, J.; Liang, Q.; et al. Automatic quantification and grading of hip bone marrow oedema in ankylosing spondylitis based on deep learning. *Mod. Rheumatol.* 2022, 32, 968–973. [CrossRef] [PubMed]
- [13] Navarini, L.; Caso, F.; Costa, L.; Currado, D.; Stola, L.; Perrotta, F.; Delfino, L.; Sperti, M.; Deriu, M.A.; Ruscitti, P.; et al. Cardiovascular risk prediction in ankylosing spondylitis: From traditional scores to machine learning assessment. *Ther.* 2020, 7, 867–882. [CrossRef] [PubMed]
- [14] Castro-Zunti, R.; Park, E.H.; Choi, Y.; Jin, G.Y.; Ko, S.-b. Early detection of ankylosing spondylitis using texture features and statistical machine learning, and deep learning, with some patient age analysis. *Comput. Med. Imaging Graph.* 2020, 82, 101718. [CrossRef]
- [15] Lin, K.Y.Y.; Peng, C.; Lee, K.H.; Chan, S.C.W.; Chung, H.Y. Deep learning algorithms for magnetic resonance imaging of inflammatory sacroiliitis in axial spondyloarthritis. *Rheumatology* 2022, 61, 4198–4206. [CrossRef]
- [16] Bressem, K.K.; Vahldiek, J.L.; Adams, L.; Niehues, S.M.; Haibel, H.; Rodriguez, V.R.; Torgutalp, M.; Protopopov, M.; Proft, F.; Rademacher, J.; et al. Deep learning for detection of radiographic sacroiliitis: Achieving expert-level performance. *Arthritis Res. Ther.* 2021, 23, 1–10. [CrossRef]
- [17] Shenkman, Y.; Qutteineh, B.; Joskowicz, L.; Szeskin, A.; Yusef, A.; Mayer, A.; Eshed, I. Automatic detection and diagnosis of sacroiliitis in CT scans as incidental findings. *Med. Image Anal.* 2019, 57, 165–175. [CrossRef]
- [18] Bressem, K.K.; Adams, L.C.; Proft, F.; Hermann, K.G.A.; Diekhoff, T.; Spiller, L.; Niehues, S.M.; Makowski, M.R.; Hamm, B.; Protopopov, M.; et al. Deep learning detects changes indicative of axial spondyloarthritis at MRI of sacroiliac joints. *Radiology* 2022, 305, 655–665. [CrossRef]