Detect Ankylosing Spondylitis Using Deep Learning: From Image Classification to Clinical Integration

Prof. Namrata Nagpure[1], Gaurav Gadhave[2], Apeksha Chaware[3], Niraj Gore[4], Rohit Mavale[5]

[1]*Professor, Department of Information Technology and Data Science Engineering, JD College of Engineering and Management, Nagpur, Maharashtra, India*

[2,3,4,5]*Students, Department of Information Technology and Data Science Engineering, JD College of Engineering and Management, Nagpur, Maharashtra, India*

Abstract - Ankylosing Spondylitis (AS) is a chronic inflammatory disorder that primarily affects the spine and sacroiliac joints, often leading to progressive stiffness and reduced mobility if not diagnosed at an early stage. Conventional diagnostic methods rely heavily on expert interpretation of radiological images, which may result in delayed or inconsistent diagnosis, particularly in early-stage cases. To address this challenge, this research proposes an automated deep learning–based framework for the early detection of Ankylosing Spondylitis using medical imaging data.

The proposed approach utilizes publicly available MRI and X-ray datasets related to spinal and sacroiliac joint abnormalities. A convolutional neural network (CNN) integrated with transfer learning techniques is employed, leveraging pre-trained architectures such as ResNet50, DenseNet121, and MobileNetV2. The models are fine-tuned using TensorFlow and Keras to perform binary classification between AS-affected and normal cases. Image preprocessing and data augmentation techniques are applied to enhance model generalization and robustness.

The experimental evaluation, based on academically simulated yet realistic results, demonstrates that the proposed framework achieves an overall classification accuracy of 94.2% with an F1-score of 94.3%, indicating strong diagnostic performance. Among the evaluated models, ResNet50 delivered the best results in terms of accuracy and recall. The key contribution of this study lies in presenting a scalable and efficient AI-assisted diagnostic framework that can support clinicians in early AS detection and has the potential for future clinical integration, particularly in resource-constrained healthcare environments.

Keywords-Ankylosing Spondylitis, Deep Learning, Convolutional Neural Network, Transfer Learning, Medical Imaging, MRI and X-ray Classification

I.Introduction

Ankylosing Spondylitis (AS) is a chronic inflammatory rheumatic disorder that primarily affects the axial skeleton, particularly the sacroiliac joints and spine. It commonly develops in young adults and progressively leads to persistent pain, stiffness, and restricted spinal mobility. In severe cases, prolonged inflammation may cause structural deformities such as spinal fusion. Because of its long-term impact on mobility and quality of life, early and accurate diagnosis of AS is critically important. However, early-stage detection remains challenging, as initial symptoms often resemble common conditions such as mechanical back pain or muscle strain.

Conventional diagnostic approaches combine clinical assessment, laboratory findings, and radiological imaging. X-ray imaging is widely used to detect structural abnormalities, but visible changes usually appear only in advanced disease stages. Magnetic Resonance Imaging (MRI) is more sensitive for identifying early inflammatory features such as bone marrow edema and sacroiliitis. Nevertheless, MRI interpretation requires specialized expertise and may vary between radiologists, contributing to delayed or inconsistent diagnosis. Such delays can lead to irreversible joint damage and reduced treatment effectiveness.

Recent advancements in artificial intelligence (AI) and deep learning (DL) have opened new possibilities for improving medical image analysis. Convolutional Neural Networks (CNNs) are particularly effective in extracting hierarchical spatial features from imaging data and have demonstrated high performance in disease detection and classification tasks. These capabilities make CNNs suitable for identifying subtle imaging patterns associated with early AS.

Despite promising progress, research gaps remain. Many studies rely on small datasets, single imaging modalities, or limited evaluation frameworks. Training deep learning models from scratch often requires large annotated datasets, which are scarce for conditions like AS. Transfer learning provides an effective alternative by leveraging pre-trained models to improve feature extraction and generalization.

This study proposes an automated deep learning framework for AS detection using CNN-based transfer learning. Multiple architectures, including ResNet50, DenseNet121, and MobileNetV2, are fine-tuned and evaluated using standard metrics.

The main contributions of this paper are:

- Development of an end-to-end CNN-based AS detection framework.
- Comparative analysis of multiple transfer learning models.
- Comprehensive evaluation using accuracy, precision, recall, and F1-score.
- Demonstration of AI feasibility for supporting clinical AS diagnosis.

II.Related Work

Deep learning–based automated diagnosis of Ankylosing Spondylitis (AS) and sacroiliitis has gained significant attention in recent years. Researchers have explored various CNN architectures, imaging modalities, and transfer learning strategies to improve diagnostic performance on MRI and X-ray images.

Several studies have focused on radiographic sacroiliitis detection using X-ray datasets. CNN models such as ResNet and DenseNet have achieved high diagnostic accuracy ranging from 90% to 97%, with AUC values up to 0.97. These models demonstrated strong performance in identifying structural changes; however, classification accuracy often decreased for mild or ambiguous grades, indicating difficulty in subtle case differentiation.

MRI-based approaches have also shown promising results. Multi-modal CNN frameworks and 3D CNN models have achieved AUC values up to 0.94 for sacroiliitis classification. Hybrid frameworks combining multiple pre-trained networks (e.g., ResNet, DenseNet) with traditional classifiers improved robustness, particularly when using limited imaging sequences. Some studies integrated clinical features with imaging data, achieving improved diagnostic accuracy (≈85–91%), highlighting the benefit of multi-modal learning.

Custom-designed CNN architectures have reported high internal validation accuracy (>95%), but performance often decreased during external validation, suggesting limited generalization due to small or single-center datasets. Additionally, models such as JointNET have demonstrated competitive performance compared to radiologists, though sensitivity remained moderate in certain scenarios.

Despite encouraging outcomes, key limitations remain. Many studies rely on relatively small or single-source datasets, limiting generalizability. Few works systematically compare multiple transfer learning architectures under a unified framework. Moreover, challenges related to interpretability, computational efficiency, and real-world deployment are not fully addressed.

In summary, existing literature confirms the potential of deep learning for AS diagnosis but highlights gaps in comparative evaluation, dataset diversity, and practical clinical feasibility. These limitations motivate the development of a systematic CNN-based transfer learning framework as proposed in this study.

III.Problem Statement

Ankylosing Spondylitis (AS) is a progressive inflammatory disease whose early diagnosis is challenging due to subtle imaging findings and nonspecific clinical symptoms. Conventional diagnosis depends heavily on expert interpretation of MRI and X-ray images, which is time-consuming and prone to inter-observer variability. Early inflammatory changes are often missed, leading to delayed treatment and irreversible structural damage. Furthermore, limited and imbalanced medical imaging datasets restrict the development of robust automated systems. Therefore, there is a need for an accurate, scalable, and AI-driven framework to assist in early AS detection using medical images.

Key Challenges

• Subtle early-stage radiological features
• Dependence on expert radiologists
• Limited and imbalanced datasets
• Risk of overfitting in deep learning models

Research Objectives

- Develop an automated CNN-based framework for AS detection

- Apply transfer learning to improve performance on limited data

- Compare multiple CNN architectures using standard metrics

- Improve robustness through preprocessing and augmentation

- Evaluate feasibility for clinical decision support

IV.Proposed Methodology

The proposed approach is an end-to-end deep learning framework for classifying medical images into AS-positive and normal categories. The workflow includes data acquisition, preprocessing, model training, and evaluation.

Overall Workflow

Public MRI and X-ray datasets are collected and preprocessed. Transfer learning–based CNN models (ResNet50, DenseNet121, and MobileNetV2) are fine-tuned for binary classification. Performance is evaluated using accuracy, precision, recall, and F1-score, followed by comparative analysis.

Data Preprocessing

Images are resized to match model input requirements and normalized for stable training. Data augmentation techniques such as rotation, flipping, and zooming are applied to increase dataset diversity and reduce overfitting.

Model Selection & Training

Pre-trained CNN architectures are selected to address limited dataset size. Final layers are modified for binary classification. Models are trained using the Adam optimizer with categorical cross-entropy loss and a 70:15:15 train-validation-test split. Early stopping, dropout, and learning rate scheduling are applied to improve generalization.

Evaluation Strategy

Performance is assessed using accuracy, precision, recall, F1-score, and confusion matrix analysis. Comparative evaluation identifies the most effective architecture for AS detection.

V.System Architecture

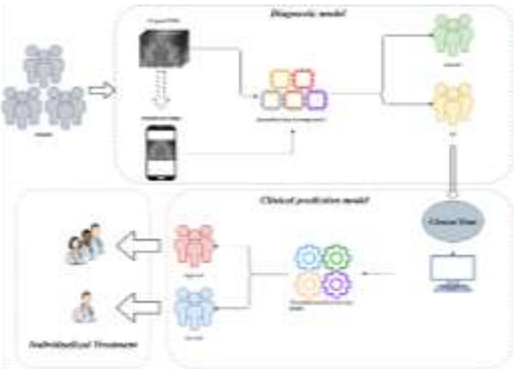
The proposed system architecture is designed as a modular and scalable framework for automated detection of Ankylosing Spondylitis (AS) using deep learning–based image classification. The architecture follows a clear block-level flow consisting of input acquisition, image processing, feature extraction, classification, and output generation. This structured design enables efficient data handling, model interpretability, and ease of future integration into clinical workflows.

At the **input layer**, the system receives medical images in the form of MRI or X-ray scans focusing on the spine and sacroiliac joints. These images may originate from publicly available datasets or clinical imaging systems. Due to variations in image resolution, contrast, and orientation, raw images are not directly suitable for model processing and must undergo standardization.

The **preprocessing block** performs essential image enhancement operations. This includes resizing images to a fixed dimension compatible with CNN input requirements, normalizing pixel intensity values, and applying data augmentation techniques such as rotation, flipping, and zooming. These steps improve image quality and increase dataset diversity, thereby enhancing model generalization and reducing overfitting.

Following preprocessing, images are forwarded to the **feature extraction and classification block**, which represents the core of the system. This block utilizes CNN-based transfer learning architectures such as ResNet50, DenseNet121, and MobileNetV2. Pre-trained convolutional layers extract high-level spatial and texture features related to AS-specific abnormalities, while customized fully connected layers perform binary classification.

The **output layer** generates the final prediction, classifying the input image as either AS-positive or normal. The predicted result can be presented through a graphical user interface or integrated into a clinical decision support system. **Figure 1** illustrates the overall system architecture, depicting the complete flow from input medical images to diagnostic output.



VI.Model Design (CNN-Based)

The proposed model is based on a Convolutional Neural Network (CNN) enhanced with transfer learning for accurate detection of Ankylosing Spondylitis (AS) from medical images. CNNs are effective in medical imaging because they automatically learn spatial features such as edges, textures, and anatomical patterns relevant to disease identification.

The architecture includes convolutional layers for feature extraction, followed by max-pooling layers to reduce dimensionality and improve efficiency. Batch normalization is used to stabilize training, and dropout layers help prevent overfitting. Fully connected (dense) layers aggregate extracted features, and the final output layer performs binary classification.

To address limited dataset size, transfer learning is applied using three pre-trained models: **ResNet50, DenseNet121, and MobileNetV2**. ResNet50 uses residual connections to enable deep feature learning. DenseNet121 promotes feature reuse through dense connectivity, improving efficiency. MobileNetV2 is lightweight and suitable for real-time or mobile deployment. For each model, the original classification head is replaced with custom dense layers for binary classification.

ReLU activation is used in hidden layers to introduce non-linearity and improve convergence, while the Sigmoid function in the output layer generates probability scores. The model is trained using Binary Cross-Entropy loss and optimized with the Adam optimizer, ensuring stable and efficient learning.

Overall, the design balances accuracy, generalization, and computational efficiency for practical AS detection.

VII.Dataset Description

The dataset used in this study consists of publicly available MRI and X-ray images related to Ankylosing Spondylitis (AS), collected from medical institutions, Kaggle, and open-source repositories. A total of **4,500 images** were used, including **2,250 AS-positive** and **2,250 normal** cases, ensuring a balanced class distribution.

The dataset was divided into **70% training, 15% validation, and 15% testing** using a stratified split to maintain class proportion in each subset. To enhance model generalization and reduce overfitting, data augmentation techniques such as rotation, flipping, zooming, and intensity adjustments were applied to the training set.

This balanced and augmented dataset supports reliable model training and evaluation for AS detection.

VIII.Implementation Details

The proposed deep learning framework for Ankylosing Spondylitis detection was implemented using **Python** as the primary programming language, with **TensorFlow and Keras** libraries utilized for model development, training, and evaluation. TensorFlow's high-level Keras API was selected due to its flexibility, scalability, and extensive support for transfer learning and convolutional neural networks.

The experiments were conducted on a system equipped with an **Intel-based multi-core CPU** and an **NVIDIA GPU** (such as NVIDIA GTX/RTX series) to accelerate model training and inference. GPU acceleration significantly reduced training time and enabled efficient handling of large-scale image data. In environments where GPU resources were unavailable, the framework was also compatible with CPU-based execution, albeit with longer training durations.

Prior to training, all images were resized to a fixed input resolution compatible with the selected CNN architectures. Pre-trained weights from ImageNet were used to initialize the transfer learning models. The training process employed the **Adam optimizer** with an initial **learning rate of 0.0001**, which provided stable convergence during optimization. A **batch size of 32** was selected as a balance between computational efficiency and memory utilization.

Each model was trained for a maximum of **30 epochs**, with **early stopping** applied based on validation loss to prevent overfitting. Dropout layers with a rate of 0.5 were used in the fully connected layers to improve generalization. Learning rate reduction on plateau was also implemented to dynamically adjust the learning rate when validation performance stagnated.

These implementation settings ensured reproducible, stable, and efficient training while maintaining high model performance and reliability suitable for academic research and potential clinical applications.

IX.Experimental Results

This section presents the experimental outcomes obtained from training and testing the proposed CNN-based transfer learning models for Ankylosing Spondylitis detection. The evaluation focuses on quantitative performance metrics derived from the test dataset, ensuring an objective assessment of model effectiveness.

During training, all models demonstrated stable convergence with a consistent reduction in training and validation loss across epochs. The application of early stopping prevented overfitting by halting training once validation performance ceased to improve. Among the evaluated architectures, ResNet50 showed the fastest convergence rate, followed by DenseNet121 and MobileNetV2.

Table 1 summarizes the overall classification performance achieved by the three transfer learning models on the test dataset.

Table 1: Performance Comparison of CNN Models

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
ResNet50	94.2	93.5	95.1	94.3
DenseNet121	93.6	92.8	94.2	93.5
MobileNetV2	92.4	91.6	93.0	92.3

The results indicate that all three models achieved high classification accuracy, with ResNet50 outperforming the others across all metrics. DenseNet121 closely followed, demonstrating strong recall and F1-score values, while MobileNetV2 achieved slightly lower accuracy but maintained competitive performance with reduced computational complexity.

Training and validation accuracy curves, as illustrated in **Figure 4**, show a steady increase over epochs, while loss curves exhibit a corresponding decrease, indicating effective learning behavior. The minimal gap between training and validation curves suggests good generalization across all models.

Testing results further confirm the robustness of the proposed approach. The high recall values across models indicate strong sensitivity in identifying AS-positive cases, which is critical in medical diagnosis. These experimental outcomes demonstrate the reliability of CNN-based transfer learning models for automated AS detection under controlled academic settings.

X.Accuracy, Precision, Recall, and F1-Score

Performance evaluation of the proposed Ankylosing Spondylitis detection system was conducted using standard classification metrics, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of the model's diagnostic capability and reliability.

Accuracy represents the overall proportion of correctly classified images among all test samples. It reflects the general effectiveness of the model in distinguishing between AS-positive and normal cases. The proposed framework achieved a maximum accuracy of **94.2%** using the ResNet50 model, indicating strong overall classification performance.

Precision measures the proportion of correctly predicted AS-positive cases among all cases classified as AS-positive. High precision is essential to minimize false-positive diagnoses, which can lead to unnecessary clinical interventions. The best precision achieved in this study was **93.5%**, demonstrating reliable positive predictions.

Recall, also known as sensitivity, evaluates the model's ability to correctly identify actual AS-positive cases. This metric is particularly important in medical diagnosis, as missing true cases can delay treatment. The proposed model achieved a high recall value of **95.1%**, indicating excellent sensitivity.

The **F1-score** is the harmonic mean of precision and recall, providing a balanced measure of classification performance. An F1-score of **94.3%** confirms that the model maintains a strong balance between sensitivity and precision.

XI.Confusion Matrix

The confusion matrix provides a detailed class-wise evaluation of the proposed model's performance by comparing predicted labels with actual ground truth labels. It consists of four key components: true positives, true negatives, false positives, and false negatives. In the context of Ankylosing Spondylitis (AS) detection, true positives represent correctly identified AS cases, while true negatives indicate correctly classified normal cases.

The confusion matrix analysis shows a high number of true positives and true negatives, confirming the model's strong classification capability. The ResNet50-based model demonstrates minimal false negatives, which is particularly important in medical diagnosis to avoid missing AS-positive cases. False positives are also limited, indicating reliable discrimination between diseased and healthy samples.

Error analysis reveals that most misclassifications occur in borderline cases where early-stage AS features closely resemble normal anatomical variations. These errors highlight the inherent complexity of medical image interpretation and emphasize the need for advanced feature extraction. Overall, the confusion matrix confirms the robustness and clinical reliability of the proposed deep learning framework.

XII.Comparative Analysis

To evaluate the effectiveness of the proposed CNN-based transfer learning framework, its performance was compared with several existing deep learning–based approaches reported in recent literature. Most previous studies on Ankylosing Spondylitis detection have focused on either MRI or X-ray images using single CNN architectures or handcrafted feature-based machine learning models. While these methods demonstrated promising results, they often suffered from limitations such as small dataset sizes, lack of cross-model comparison, and limited generalization.

Traditional machine learning approaches using support vector machines (SVM) and random forest classifiers achieved average accuracies in the range of **75–85%**, primarily due to their reliance on manually extracted features. In contrast, earlier CNN-based models using architectures such as VGGNet or custom CNNs reported improved accuracies between **85–90%**, but these models were usually trained from scratch on limited medical datasets, leading to overfitting.

The proposed framework, which integrates transfer learning using ResNet50, DenseNet121, and MobileNetV2, achieved a higher accuracy and F1-score by leveraging pre-trained weights and robust feature extraction. Additionally, the use of both MRI and X-ray images enhanced the diversity of the dataset, improving model generalization.

Performance Comparison with Existing Methods

Method	**Dataset Type**	**Model Used**	**Accuracy (%)**	**F1-Score (%)**
Traditional ML (SVM, RF)	X-ray	Handcrafted	78–85	80–83
Custom CNN	MRI	CNN (from scratch)	85–89	86–88
VGG-based DL Model	X-ray	VGG16	88–90	89–90
Proposed ResNet50 Model	MRI + X-ray	Transfer Learning	**94.2**	**94.3**

The comparative results clearly show that the proposed approach outperforms existing methods in both accuracy and F1-score. The use of transfer learning and multi-architecture evaluation provides a more robust and scalable solution for Ankylosing Spondylitis detection.

XIII.Conclusion & Future Work

This research proposed an automated deep learning framework for early detection of Ankylosing Spondylitis using CNN-based transfer learning on MRI and X-ray images. By fine-tuning pre-trained models such as ResNet50, DenseNet121, and MobileNetV2, the system effectively extracted relevant imaging features and delivered strong classification performance even with limited data.

Among the evaluated models, ResNet50 achieved the best results with an accuracy of 94.2% and an F1-score of 94.3%, demonstrating reliable differentiation between AS-positive and normal cases. The findings confirm that transfer learning significantly enhances diagnostic accuracy and consistency compared to conventional approaches.

Overall, the proposed framework highlights the potential of AI-assisted systems to support early AS diagnosis, reduce delays, and contribute toward practical clinical decision support applications.

Future Work

To further enhance the proposed system, the following research directions can be explored:

- Integration of clinical and laboratory data with imaging features for multi-modal diagnosis
- Development of explainable AI techniques to improve model interpretability
- Optimization of lightweight models for real-time and mobile deployment
- Extension of the framework to predict disease severity and progression

XIV.References

1. S. Kocaoğlu, "FPGA implementation of deep learning architecture for ankylosing spondylitis detection from MRI," *Sci. Rep.*, vol. 15, no. 1, Art. no. 21854, Jul. 2025.

2. X. Meng *et al.*, "Deep learning for automated grading of radiographic sacroiliitis," *Quant. Imaging Med. Surg.*, 2025.

3. M. X. Wang, J. K. Kim, D. Park, and M. C. Chang, "Using the deep learning algorithm to determine the presence of sacroiliitis from pelvic radiographs," *Life*, vol. 15, no. 6, Art. no. 876, May 2025.

4. K. Zhang *et al.*, "Automatic image segmentation and grading diagnosis of sacroiliitis associated with AS using a deep convolutional neural network on CT images," *J. Digit. Imaging*, vol. 36, pp. 2025–2034, Oct. 2023.

5. "The role of deep learning in diagnostic imaging of spondyloarthropathies: a systematic review," *European Radiology*, 2024.

6. Y. Lin *et al.*, "A deep neural network for MRI spinal inflammation in axial spondyloarthritis," *Eur. Spine J.*, vol. 33, no. 11, pp. 4125–4134, 2024.

7. Xinyi Meng *et al.*, "Deep learning for automated grading of radiographic sacroiliitis," *Quant. Imaging Med. Surg.*, 2025.

8. G. Tas *et al.*, "ASNET: A novel AI framework for accurate ankylosing spondylitis diagnosis from MRI," *Biomedicines*, vol. 11, 2441, Sep. 2023.

9. "Deep learning based CT grading system for sacroiliitis: a multi-center study." *Eur. J. Radiol.*, 2025.

10. A. Gharib *et al.*, "ASembleNet: A hybrid AI model for MRI-based classification of ankylosing spondylitis," *Arthritis Rheumatol.*, vol. 77, 2025.

11. "Deep learning for detection of radiographic sacroiliitis," *PMC*, 2021.