International Journal of Innovations in Science & Technology

Computer-Aided System for the Detection of Rheumatoid Arthritis

Aqib Fayyaz Shaheer*, Mubbashir Ayub Minhas
Department of Software Engineering (UET Taxila, Punjab, Pakistan).
***Correspondence**: 17-se-97@alumni.uettaxila.edu.pk

Citation | Shaheer. A. F, Minhas. A. M, "Computer-Aided System for the Detection of Rheumatoid Arthritis", IJIST, Vol. 07 Issue. 03 pp 1630-1645, July 2025
DOI | https://doi.org/10.33411/ijist/20257316301645
Received | June 10, 2025 **Revised |** July 22, 2025 **Accepted |** July 23, 2025 **Published |** July 24, 2025.

Rheumatoid Arthritis (RA) is a chronic disease that causes disability in movement. RA classification is critical for effective diagnosis and treatment planning. This work explores the application of the EfficientNetB6 architecture using transfer learning to classify RA severity into three categories: Healthy, Moderate, and Severe. Medical imaging dataset containing X-Ray images, enhanced with contrast-limited adaptive histogram equalization (CLAHE), data augmentation techniques, and fine-tuning of hyperparameters was applied in this work. We compared EfficientNetB6 with all the models of the EfficientNet family and all other state-of-the-art models. When we combined EfficientNetB6 with CLAHE, we achieved the highest accuracy of 96.06%. Without using CLAHE, accuracy dropped by 4% to 5% for all the models. For a healthy class model, we achieved precision, recall, and F1-score of 99.36%,97.81%,98.58% respectively, showing robustness in identifying healthy cases. Moderate class yielded precision, recall, and F1-score of 89.45%,95.07%,92.17% respectively, demonstrating the model's effectiveness in identifying moderate cases with minimal false negatives. The Severe class presented more challenges with a precision, recall, and F1-score of 85.11%,78.43%, 81.63% highlighting the need for improved recall value. To further improve results, we suggest enhancements such as advanced data augmentation and synthetic data generation, particularly for the Severe class, consequently aiding clinicians in the identification of RA.

Keywords. Rheumatoid arthritis, Deep learning, Disease detection, Medical Imaging.

Introduction:

Rheumatoid Arthritis (RA) is a long-term autoimmune disorder characterized by persistent inflammation and progressive damage to the joints, resulting in pain, joint stiffness, and reduced mobility [1]. The knee is one of the joints most commonly affected by Rheumatoid Arthritis (RA), and it can suffer severe damage if the condition is not identified and managed promptly. Traditional diagnostic methods for detecting RA in the knee generally involve clinical assessments alongside imaging techniques such as X-rays and magnetic resonance imaging (MRI) [2]. However, these methods often face challenges related to accuracy and cost-effectiveness. As a result, there is a pressing need for a more efficient and accurate method to detect RA in the knee. Recently, deep learning techniques for medical image analysis have garnered significant attention. Models like Convolutional Neural Networks (CNNs) have demonstrated promising outcomes in various medical imaging tasks, such as detecting and classifying diseases like breast cancer [3], lung cancer [4], and diabetic retinopathy [5]. However, to the best of our knowledge, there is limited research focused on the application of deep learning techniques for detecting Rheumatoid Arthritis (RA) specifically in the knee joint. Therefore, the proposed study aims to address this gap in existing literature by investigating the feasibility of utilizing deep learning techniques for the detection of Rheumatoid Arthritis (RA) in the knee joint. In Rheumatoid Arthritis (RA) patients, the synovial membrane surrounding the knee joint becomes inflamed, leading to the breakdown of cartilage and bone in the joint.

This inflammation can also lead to the formation of scar tissue and the accumulation of excess fluid in the joint, resulting in pain, stiffness, and limited range of motion. Over time, the cartilage and bone damage can cause deformities in the knee joint and surrounding structures, which can further exacerbate pain and functional impairment. Effective treatment of RA in the knee typically involves a combination of medications, physical therapy, and lifestyle modifications. In some cases, surgery may be necessary to repair or replace damaged joint structures. Early diagnosis and treatment are crucial in preventing joint damage and improving long-term outcomes for RA patients.

Literature Review:

Diagnosing RA, particularly in its early stages, remains a complex and challenging task due to subtle clinical signs and overlapping symptoms with other joint disorders. A variety of methods, including deep convolutional neural networks (CNNs), machine learning, and deep learning algorithms, have been proposed to tackle various issues related to Rheumatoid Arthritis data. In clinical research, magnetic resonance imaging (MRI) serves as a primary tool for diagnosing Rheumatoid Arthritis and is utilized for tasks such as feature extraction [6]. Before the rise of deep learning, machine learning techniques have been employed since 2000, often paired with manually engineered features, necessitating the expertise of domain specialists to achieve high accuracy and optimal performance. Since 2013, new architectures featuring deeper models have become increasingly popular, particularly in the realm of medical image processing [7]. Numerous deep learning algorithms have been employed to detect Rheumatoid Arthritis in patients, including the use of support vector machines on weighted MRI images and random forest classifiers for multimodal classification [8].

Shamir et al. [9] present a computer-aided diagnostic (CAD) system designed to detect osteoarthritis (OA) in knee X-ray images. Their approach initially extracts features using discrete wavelet transform (DWT) and gray level co-occurrence matrix (GLCM) techniques. These features are subsequently classified with a support vector machine (SVM) to ascertain if the knee joint is normal or impacted by OA. The proposed method achieves a high accuracy rate of 93.75% in detecting OA in knee X-ray images. However, one limitation of this method is that it only focuses on detecting OA in the knee joint, and no other types of arthritis that can affect other joints in the body. Additionally, the proposed method relies on the accuracy of the feature extraction techniques and the SVM classifier, which may not perform optimally in cases where the X-ray images are of poor quality or contain artifacts. Zeng et al. [10] provided a comprehensive review of studies that have used deep learning approaches for discovering imaging features related to the diagnosis and prognosis of knee osteoarthritis (OA). Zeng et al. first provided an overview of knee OA disease and its diagnostic imaging modalities, such as X-ray, MRI, and CT scans. They then reviewed various manual imaging grading systems for knee OA, including Kellgren- Lawrence (K-L) grading and OARSI Atlas grading, and discussed their limitations in terms of reproducibility and sensitivity. Subsequently, the authors explored machine learning- based approaches for discovering knee OA imaging features using computer vision and deep learning techniques. They discussed various studies that used machine learning for automated knee OA diagnosis and prediction, such as the use of convolutional neural networks (CNNs) and support vector machines (SVMs).

Zeng et al. concluded that while manual imaging grading systems are widely used in clinical settings, they are subjective and prone to inter-observer variability. On the other hand, machine learning techniques have shown good results in detection of image features and classification of Knee OA. For instance, Alexopoulos et al. [11] employed a deep learning approach using convolutional neural networks (CNNs) for the early detection of knee OA. Their model utilized CNNs to extract relevant image features and was trained on a large dataset of MRI scans, achieving an accuracy of 90.5% in distinguishing healthy knees from those affected by OA. This study has some limitations which included not using the new images but instead using the already available images that were quite old, and they did not consider the broader perspective but specific population. Their adaptability to the results remains unknown to other populations. Alexopoulos et al. also acknowledge that the interpretability of their system is very limited in identifying which image feature contributed to the system's classification decision. Guan et al. [12] focused on identifying the severity of Knee OA. Instead of relying on subjective interpretation of Knee Images, they used a deep learning approach where they trained a deep CNN model on Knee radiographs consisting of Knee OA severity. Their findings suggest that the proposed method outperforms traditional techniques in terms of both accuracy and overall performance. However, their approach has certain limitations, which include evaluating the performance of their CNN model using only radiographs but not considering the other image modalities to see the broader perspective in identifying the severity of knee OA. Also, their study was limited by the use of a relatively small dataset that did not represent a broader population. Additionally, their approach focused exclusively on knee OA, without considering other joints that can also be significantly affected by osteoarthritis.

Murakami et al. [13] proposed a new approach to identify bone erosion from Rheumatoid Arthritis using hand radiographs with convolutional neural networks. They used a dataset consisting of 825 hand radiographs with a two-stage training strategy to improve the classification performance of their model. They obtained specificity of 95.6% and sensitivity of 93.8% in detection of erosion caused by Rheumatoid Arthritis. Although they achieved overall promising results, their study had several limitations. It relied solely on hand radiographs, excluding other imaging modalities such as MRI or X-ray, which could provide more comprehensive insights. Furthermore, the study did not incorporate critical diagnostic features such as joint space narrowing, which play a significant role in the detection of Rheumatoid Arthritis. Another notable drawback of their study is the lack of comparison with other state-of-the-art methods, which limits the ability to evaluate the true effectiveness and competitiveness of their proposed approach. Wang et al. [14] used a deep learning approach for detection of Rheumatoid Arthritis in hand X-ray images. They used 7000 X-ray images for training the model and then used another dataset consisting of 156 images for testing the model, which limits the generalizability of results. Their system attained a specificity and sensitivity of

92.9% and 87.5% respectively, with a correlation coefficient of 0.797. Their approach showed promising results in identifying joint space narrowing and erosion. Despite its promising results, the approach had several limitations, including—but not limited to—the exclusion of important patient data such as demographics and disease activity. Additionally, the study relied solely on hand X-ray images, without incorporating other imaging modalities that could enhance diagnostic accuracy. They used a small dataset consisting of 156 images only and not including the greater population.

Salmeron et al. [15] proposed a novel-based approach using Fuzzy C-Means (FCM) clustering and Particle Swarm Optimization (PSO) for detection of Rheumatoid Arthritis. The system uses a limited number of features to identify if a patient has Rheumatoid Arthritis (RA) positive or not, but it uses a very limited number of input images that include 50 images from RA-positive cases and 50 from healthy cases, which is very low for machine learning algorithms. In terms of accuracy, specificity, sensitivity, and F1-Score, their system outperformed traditional methods but with some limitations that include using very small data for training and testing. Another major drawback is that their process requires manual feature selection before applying PSOFCM for classification, which is subjective and prone to errors. Lastly, they did not compare their approach with other machine learning or deep learning methods to find out the true potential of their method. Dziekan et al. [16] used Fluorescence Time Correlation Spectroscopy (TCS) for detection of RA, which is non-invasive method to measure the mobility of fluorescence label molecules in cells. This method helps to find the mobility of synovial fluid in joints of patients suffering from RA. The study shows a significant difference in mobility for RA patients than those of healthy individuals, showing that it can be used for diagnosis of RA. Despite the fact that it can be used for early detection of RA, this study has few limitations. First, they used a very small dataset where they used 15 samples from RA patients and 10 healthy individuals, affecting generalizability of their results. Second, they did not compare the TCS approach with other diagnostic methods, including X-ray and MRI, raising questions mark of how this approach compares with other state-of-the-art methods available.

Ebert et al. [17] conducted a study that explored the use of a fluorescent imaging agent designed to bind specifically to inflamed synovial tissue in patients with Rheumatoid Arthritis (RA). This approach aimed to enhance the visualization of inflammation, potentially improving the accuracy of RA detection and assessment. This is non-invasive and injected into a vein. Then, fluorescence imaging system is analyzed to check the inflamed tissue, indicating the presence of RA in the patient under observation. The imaging agent they used is called V-pyridoxyl-5-methyltryptophan (V-PMT). In this study, they used 10 RA patients and obtained sensitivity of 90% and specificity of 100%. The authors also showed that this system is able to identify RA in patients where traditional methods like X-ray and MRI showed false negative results. This approach has some major drawbacks, including insertion of non-invasive imaging agent into patient veins and then analyzing that agent with a specific imaging analysis technique, which requires experts and special equipment. The dataset used is very small, which requires further analysis to find the generalizability and robustness of the proposed approach.

Frize et al. [18] employed a non-invasive technique for the detection of Rheumatoid Arthritis (RA) using infrared imaging. This method aimed to identify temperature variations associated with inflammation, offering a radiation-free alternative for assessing RA-related joint activity. By analyzing the temperature distribution of joints, authors claimed infrared imaging technique can help to early detect the presence of RA early in patients. Their data includes 28 RA patients and 28 healthy subjects. They used thermal cameras to capture the infrared images and then used statistical methods to find the temperature distribution of joints. Results showed the temperature distribution of joints for RA patients was significantly different than that of healthy subjects. The authors also used a conventional machine learning approach using Support Vector Machine to classify infrared images either as RA positive or as healthy individuals with an accuracy of 89.3%. Their study has several limitations, including using a very small amount of data, which affects the adaptability of results. They only focused on knee joints and not the other joints that can be affected by RA. The most important one is that their study focused only on temperature distribution, which can easily be affected by certain conditions like skin temperature and surrounding temperature. Therefore, further research is necessary to assess the effectiveness of the proposed method in a larger sample size, for various joints, and under different environmental conditions. After doing an extensive literature review, the proposed study selected Kumar and Goswami [19] as the base paper as it has state-of-the-art accuracy of 91.03%, % highest amongst all the existing approaches based on deep learning, used same evaluation parameters, state-of-the-art image processing techniques and recent publication which made it more suitable for comparison with our work.

The rest of the manuscript is structured as follows: The objectives and novelty section contain the objectives and novelty statement. Materials and Methods presents our research methodology and tools utilized. Results and Discussions contain experimental results and performance comparison.

Objectives and Novelty:

The main purpose of this study is to develop a robust and accurate deep learning-based classification system for early detection and prevention of Rheumatoid Arthritis (RA) in knee joint radiographs, using advanced image preprocessing techniques and state-of-the-art convolutional neural network architectures.

Contributions of this research are listed as follows:

We fine-tuned a robust deep learning model, EfficientNetB6, using state-of-the-art Image processing techniques (CLAHE) for accurate classification of Rheumatoid Arthritis in Knee.

We performed the balancing and augmentation of data using Image Data Augmentation to mitigate bias and enhance the generalizability of results.

We compared the performance of selected models with other state-of-the-art models in family, which we also trained on the same dataset for identification of robust solutions.

We developed an efficient tool for rheumatologists and radiologists for accurate diagnosis of Rheumatoid Arthritis in knee joint radiographs.

We fine-tuned a deep learning model, EfficientNetB6, on the OAI dataset, consisting of X-ray images of Knee with KL grade assigned to each image, utilizing Contrast Limited Adaptive Histogram Equalization, for accurate diagnosis of Rheumatoid Arthritis in knee Joint Radiographs. CLAHE works by enhancing contrast and reducing noise amplification while maintaining the global structure of the image. CLAHE helped preserve bone texture information and joint space visibility while maintaining diagnostic quality. By incorporating CLAHE in our data preprocessing step and using an advanced deep learning model, our approach aims to improve classification performance for early detection and prevention of RA in Knee radiographs.

Material and Methods:

This section discusses details of the working mechanism for the detection of Rheumatoid Arthritis using EfficientNetB6. Furthermore, the working mechanism consists of three main steps: data preprocessing, model training, and validation. We performed class balancing, data augmentation, contrast-limited adaptive histogram equalization (CLAHE), and color changing of black and white into RGB for a deep learning model. Next, with the help of transfer learning, we used the EfficientB6 model trained on ImageNet data, fine-tuned the EfficientNetB6 model, and finally, validated the model for accurate detection of Rheumatoid Arthritis. The details of the preprocessing and the architecture of the proposed system are discussed in the following sections. Figure. 1 displays the working mechanism of this work, while Figure. 2 shows the flow diagram of the methodology.

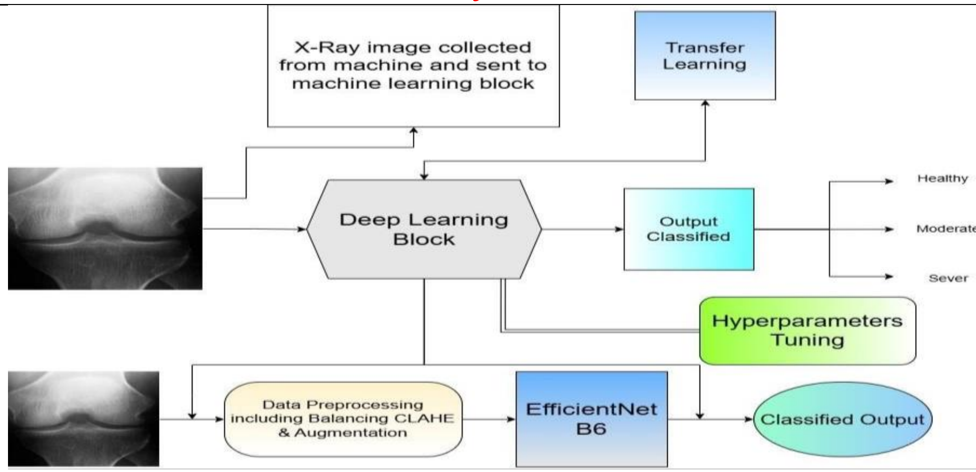

Figure 1. Working mechanism.

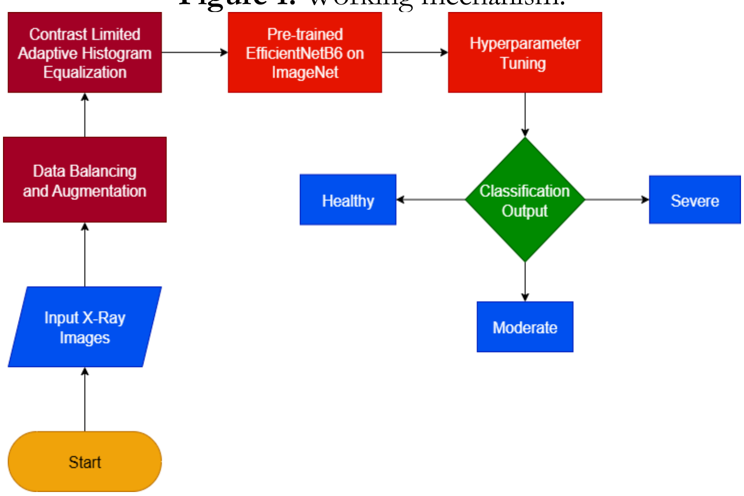

Figure 2. Flow Diagram of methodology.

Data Description:

The dataset used in this study was obtained from the Osteoarthritis Initiative (OAI), a large-scale project aimed at advancing research in osteoarthritis. It includes X-ray images of both the left and right knees; each annotated with Kellgren-Lawrence (KL) grades to indicate the severity of osteoarthritis. The images were in different dimensions ranging from 384x384 pixels to 512x512 pixels. Images were taken from approximately 4796 subjects. The images contained all the factors, including coronal, sagittal, and axial views, which provided a clear picture of joint cartilage of the knee, consequently aiding in early diagnosis and detection of disease [23]. The original dataset consisted of five classes, Grade 0 (Healthy), Grade 1 (Minimal), Grade 2 (Doubtful), Grade 3 (Moderate), and Grade 4 (Severe). In our work, we consolidated the three classes (Minimal, Doubtful, and Moderate) into one class, Moderate, as these three classes (Minimal, Doubtful, and Moderate) had overlapping radiographic features, which included inflammatory properties, progressive joint space narrowing, and erosive pattern changes. In this study, Rheumatoid Arthritis (RA) was classified into three categories—Healthy, Moderate, and Severe—to reflect the progression of the disease and support appropriate clinical decision-making. This categorization enabled more targeted management strategies: Healthy cases did not require intervention, Moderate cases were considered suitable for medical treatment to manage inflammation and prevent progression, and Severe cases often required more intensive therapies, including surgical intervention. Besides just image data, the dataset also contained metadata for each subject like their age, gender, race, and clinical data, which included subject physical examination, body index, and pain scores. This comprehensive collection of images and clinical data has been an invaluable source for researchers and scientists in overcoming the progression of osteoarthritis and in early detection and diagnosis by developing predictive models using deep learning and machine learning approaches.

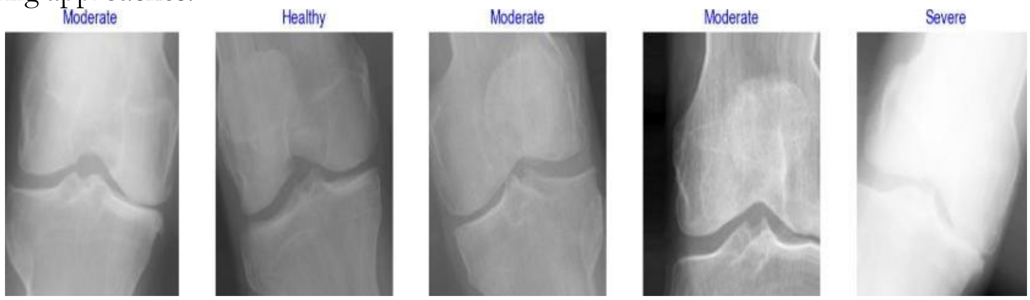

Figure 3. Images from Knee Osteoarthritis Dataset.

Data Balancing using Trimming and Augmentation:

Balancing the dataset forces the model to pay equal attention to all classes, thereby enhancing its ability to make accurate predictions across the board. The original dataset had significant imbalance with 2286 healthy, 757 moderate, and only 173 severe x-ray images, respectively. If we fed this directly to our model, we would get biased results; the model's performance and accuracy would be biased towards the healthy class, thus giving false results. To address these class imbalance issues, first, we trimmed all the classes to have a maximum of 1000 samples and a minimum of 173 samples, as severe class only contains 173 samples. Then we used image data augmentation on classes that had samples less than 1000 (Moderate and Severe). We used ImageDataGenerator from the Keras library for real-time data augmentation, and parameters we used were rotation, height/width shift, horizontal flip, and zoom. We generated 263 images for moderate class and 827 for severe class. After trimming and applying data augmentation techniques, we obtained 1,000 images for each class—Healthy, Moderate, and Severe. This process not only ensured class balance but also introduced variability through different image transformations, thereby improving the model's robustness and its ability to generalize across diverse data.

Image enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE):

After applying data balancing using augmentation, conversion of images to RGB, image enhancement using OpenCV library, and sharpening using different kernel values, we applied different models, and with EfficientNetB6, we were able to achieve the highest accuracy of 91.13 %, more than our base paper [19], which is 91.02%. This minor improvement led towards further experimentations. We analyzed that X-ray images in our dataset had varying contrast and noise, due to which the model faced difficulties in learning all the patterns in the data, and consequently, our results were not getting much better [24]. Following the evaluation of various data preprocessing methods and the optimization of hyperparameters, we applied CLAHE (Contrast Limited Adaptive Histogram Equalization) to our dataset to enhance image contrast and feature visibility. CLAHE is an extension of the traditional histogram equalization technique where all the intensity values of images are evenly distributed across the entire histogram. It works by operating on small tiles in an image, making each detail in a small region more visible. Traditional Histogram Equalization methods work by applying enhancement globally on images, resulting in over-enhancement and loss of local details. Adaptive Histogram Equalization causes over-amplification of noise and can create artificial edges, and is computationally intensive. CLAHE helps by limiting the contrast and helping alleviate the noise amplification. CLAHE preserves bone texture information and joint space visibility, maintaining diagnostic quality. While experimenting with CLAHE, we used different values of clip size and grid size. We got the highest accuracy of 96.06% when we used a clip size of

2.0 and a grid size of 8x8 with EfficientNetB6. Upon trying higher values of clip size, we noticed over-amplification of noise and loss of important radiographic features in images, resulting in degradation in performance of models. After experimenting with different grid sizes, we found that a grid size of 8x8 provides the balance between maintaining the global image characteristics and enhancing the local image features. By enhancing the contrast, the bones in X-ray images became more visible and clearer, thus helping the model to better capture the visual features and enhancing the learning pattern of joint identification.

Figure 4. Images from original dataset without applying CALHE

Figure 5. Shows enhanced images after applying CALHE.

System Architecture:

For the model selection phase, we experimented with several state-of-the-art image classification models, including InceptionResNetV2—as used in our base paper—as well as all variants of the EfficientNet family. Our results showed that the EfficientNet models outperformed the others, largely due to their advanced architecture, which offers a balanced trade-off between accuracy and computational efficiency. Figure 6 shows the image taken from the official EfficientNet paper.

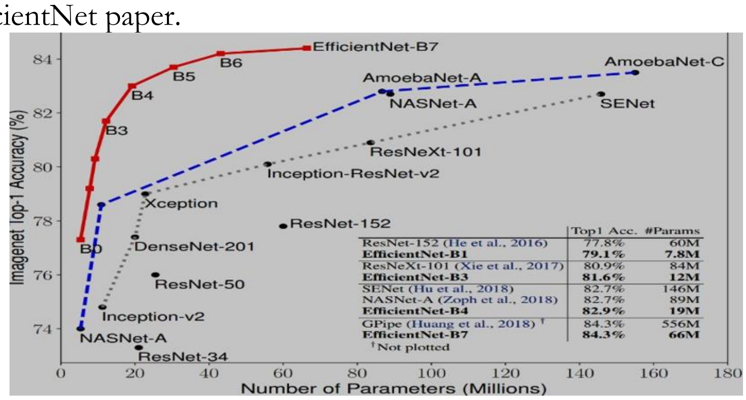

Figure 6. Model size vs ImageNet Accuracy.

In the image, the x-axis shows accuracy of models on ImageNet dataset, and the y-axis represents the number of parameters (size) of each model. As we can see, the EfficientNet family outperformed all the other state-of-the-art models. EfficientNetB7 achieved the highest accuracy of 84.3% amongst all on ImageNet, with 8.4 times smaller and 6.1 times faster than GPipe. One of the key reasons for selecting the EfficientNet family in our transfer learning approach was its ability to effectively balance model size, computational efficiency, accuracy, and feature extraction. These strengths made it well-suited for fine-tuning on our dataset, resulting in a more reliable and robust classification performance. EfficientNet family's models are based on compound scaling techniques, which uniformly scale the network width, depth, and resolution based on scaling coefficients. Because of this uniform scaling and balance between all the dimensions, they outperformed other models in computational efficiency and performance. We initialized the EfficientNetB6 model pretrained on ImageNet data and removed its top layer, and added a max pooling layer for down-sampling. Unlike the common practice, we kept the base model trainable to achieve better results. EfficientNetB6 architecture is divided into different components, including compound scaling, Mobile inverted bottleneck convolution blocks, convolution blocks, batch normalization, L1 and L2 regularizations, and dense layers. Compound scaling is represented in equation 1.

$$\text{Width} = \beta^{w}, \text{Depth} = \alpha^{d}\ , \text{Resolution} = \gamma^{r} \quad (1)$$

Where w, d, and r represent the scaling values for width, depth, and resolution, respectively, and α, β, and γ are the constants. EfficientNetB6 uses the same baseline architecture as EfficientNetB0 but with larger dimensions. Next, EfficientNetB6 uses a series of convolutions to extract features from the input images. In addition, MBConv Blocks are Mobile Inverted Bottleneck Convolutional blocks, which are highly efficient in terms of computational cost and memory usage. Batch normalization layers were incorporated following each convolutional layer to normalize the inputs of each layer. This helped in stabilizing and speeding up the training process.

$$BN\,(x) = \gamma\,((x - U / \sqrt{\sigma^2} + \varepsilon) + \beta) \qquad (2)$$

where μ and σ^2 represent the mean and variance of the input, respectively, and ϵ is a small constant to avoid division by zero. Additionally, γ and β are trainable parameters. EfficientNetB6 primarily uses the Swish activation function, which improves performance over $ReLU$ in many cases

$$\text{Swish}(x) = x.\sigma(x) \ =^{x}$$
$$1 + e^{-x} \ (3)$$

For pooling, the architecture utilized global average pooling to decrease the spatial dimensions of the feature maps before the final dense layers. We performed batch normalization on the output of the base model, which helped in stabilizing and speeding up the training process. This was followed by a dense layer with 256 units, regularized using L2 regularization (l2 = 0.016) and L1 activity and bias regularization (l1 = 0.006), and activated using the Swish function. A dropout layer with a dropout rate of 0.4 was used to mitigate overfitting, randomly deactivating a portion of input units during training by setting them to zero. The final output layer used a $softmax$ activation function to produce probability distributions over the class labels, which made it suitable for multi-class classification. For optimizer, we used Adamax with a learning rate of 0.001. The loss function chosen is categorical cross-entropy, suitable for multi-class classification problems, and for evaluation metric, we used accuracy. In addition, L1 regularization, L2 regularization, and dropout layer are represented by equations 4,5, and 6, respectively. Specifically, the L2 regularization is represented by Eq. 4

$$L2(w) = \lambda \sum w^2 \ (4)$$

where λ is the regularization strength and w^2 are the weights. Equation 5 represents the L1 regularization.

$$L1(w) = \lambda \sum |wi| \quad (5)$$

Where w represents the weights and λ represents the regularization strength. Equation 6 represents the dropout layer.

$$0\ with\ probabiity\ p$$
$$Dropout\ (x) = \{x/1 - p\ eith\ probability\ 1 - p \quad (6)$$

In the above equation, p represents the dropout rate. By using data augmentation for balancing the data and CALHE for enhancing the images and then feeding it to a sophisticated neural network, opted with appropriate regularization and hyper-parameter optimization, our classification model was able to perform accurate and robust classification for early detection of Rheumatoid Arthritis.

Results and Discussion.

This section highlights the performance outcomes of all EfficientNet family models. We created Table 1 listing the results of all EfficientNet family models that we used for experimentation with CLAHE applied. For EfficientNetB6, we provided separate detailed tables (Tables 2 and 3) using key metrics such as accuracy, precision, recall, and F1-score, both with and without using CLAHE.

Performance on all EfficientNet Models:

Table 1. EfficientNet Classification Results Summary

Model	Total Accuracy
EfficientNetB0	93.32%
EfficientNetB1	87.26%
EfficientNetB2	93.98%
EfficientNetB3	93.76%
EfficientNetB4	*94.90%*
EfficientNetB5	93.76%
EfficientNetB6	96.06%
EfficientNetB7	94.85%

Performance on EfficientNetB6:

Table 2. Classification Report on EfficientNetB6 with CLAHE (Accuracy 96.06**%)**

	precision	recall	F1-score	support
Healthy	0.9936	0.9781	0.9858	639
Moderate	0.8949	0.9507	0.9217	223
Severe	0.8511	0.8843	0.8163	51
Macro Average	0.9131	0.9044	0.908	913
Weighted Average	0.9615	0.9606	0.9607	913

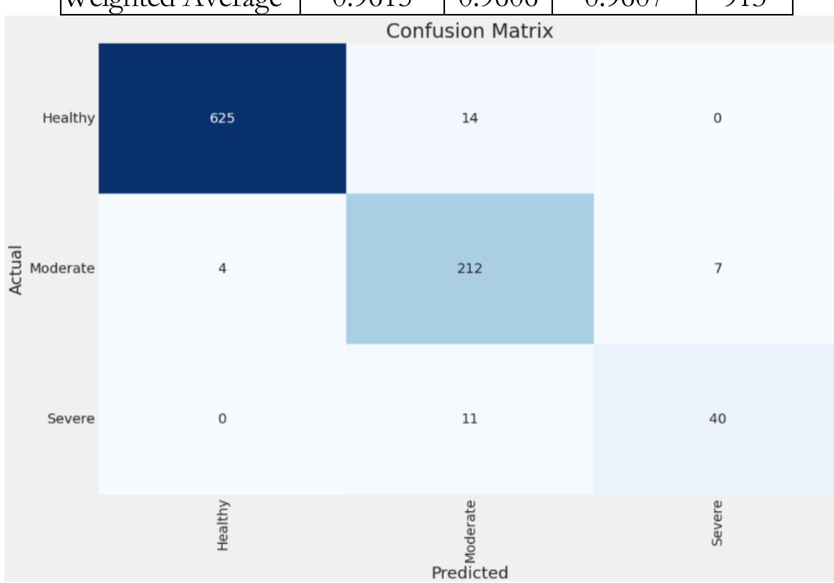

Figure 7. Confusion Matrix for EfficientNetB6 with CLAHE.

Table 3. Classification Report on EfficientNetB6 without CLAHE Accuracy (91.13%)

	precision	recall	F1-score	support
Healthy	0.9900	0.9280	0.9580	639
Moderate	0.8863	0.8386	0.8618	223
Severe	0.4330	0.8235	0.5676	51
Macro Average	0.7769	0.8783	0.8070	913
Weighted Average	0.9388	0.9113	0.9209	913

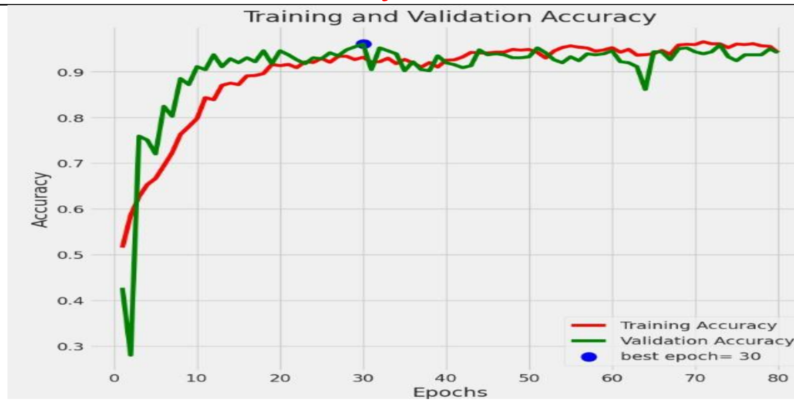

Figure 8. Training and Validation accuracy on EfficientNetB6 with CLAHE.

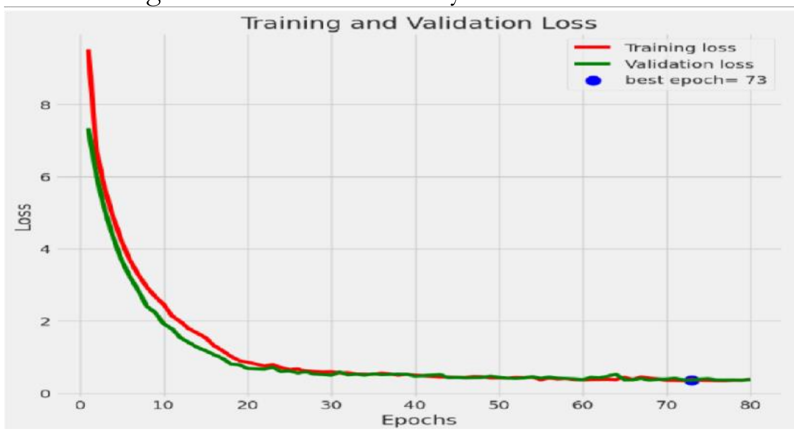

Figure 9. Training and Validation loss on EfficientB6 with CLAHE.

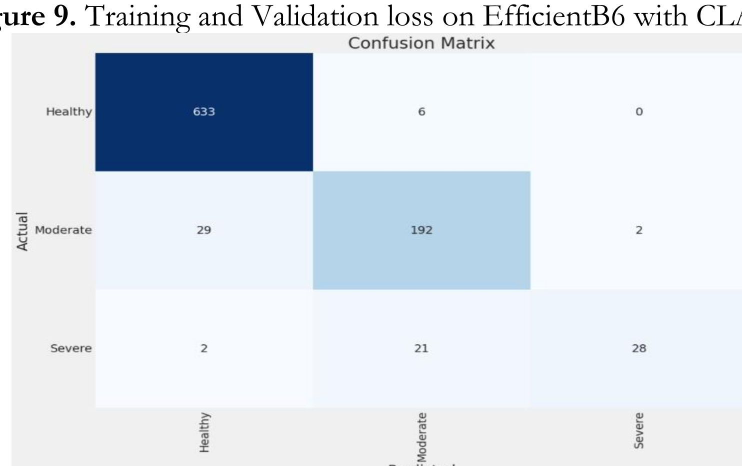

Figure 10. Confusion Matrix for EfficientNetB6 without CLAHE.

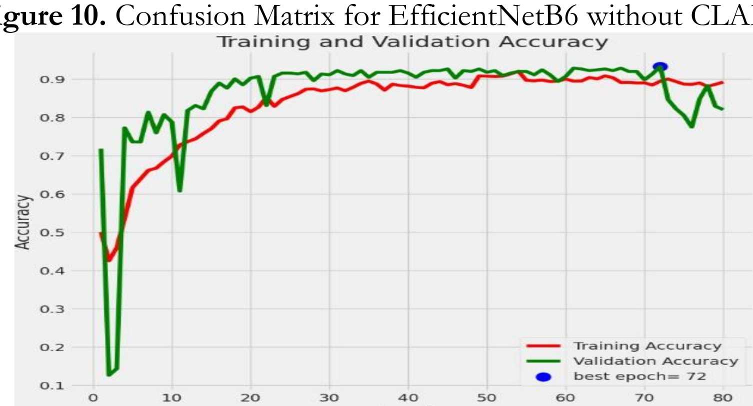

Figure 11. Training and Validation accuracy on EfficientNetB6 without CLAHE

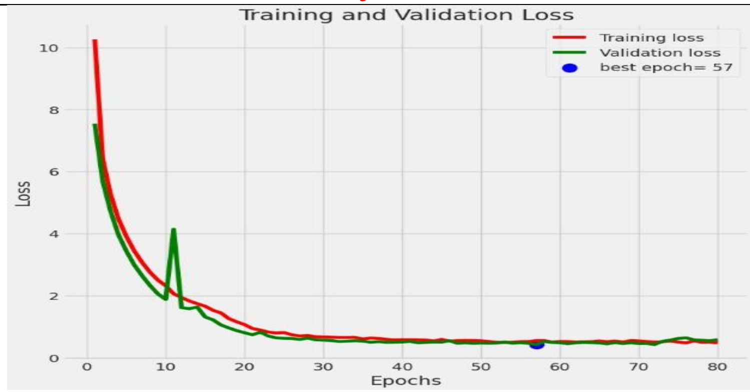

Figure 12. Training and Validation loss on EfficientNetB6 without CLAHE.

Performance Comparison:

Table 1 and graph in Figure 13 collectively illustrate the effectiveness of each model in terms of overall accuracy. Table 2 and Table 3 with detailed values of all evaluation metrics show that EfficientNetB6 outperformed all other models, achieving the highest scores across all evaluation metrics. Figure 14 presents a comparative analysis of the various approaches used for classifying Rheumatoid Arthritis, with particular emphasis on the superior performance of EfficientNetB6. The graph in Figure. 14 shows that our approach with EfficientNetB6 achieved the highest accuracy of 96.06% higher than our base paper, which is 91.02%. All the accuracy values in the graph range from 63.04% to 91.02% which are comparatively lower than our approach. Which clearly states that EfficientNetB6, due to its compound scaling method coupled with optimized hyperparameter tuning, effective use of transfer learning, robust regularization, and enhanced image processing technique using CLAHE, delivers higher performance.

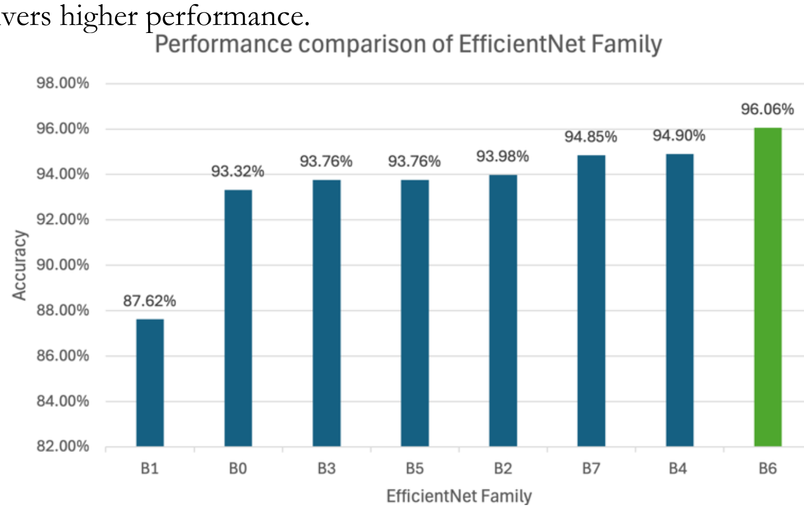

Figure 13. Performance Comparison amongst all EfficientNet Models.

Discussion and Analysis:

As shown on the graph in Figure 14, Antony et.al [12] achieved a lowest accuracy of 63.4% compare to all other studies, This could be due to several factors including use of less advanced image preprocessing techniques, not using sophisticated model architecture, using smaller dataset and using traditional machine learning approaches which were best at their time but now has been surpassed by more advanced architectures like EfficientNet family. Chen et.al [20] achieved slightly Higher accuracy of 67.71% compared to Antony et.al due to certain improvements, including but not limited to adaptation of more advanced image processing techniques and using deeper neural network architecture. Lim et.al [21] achieved an accuracy of 75.82, higher than Antony et.al and Chen et.al, but not higher than our base paper and Efficient Net Family. From all the past studies on RA detection, Tiulpan et.al [22] achieved the highest accuracy of 80% due to the use of a more robust approach, potentially incorporating convolution neural networks (CNNs). Their results were still significantly lower than our base paper, which indicates the use of less optimized architecture, not using data augmentation, and insufficient regularization techniques. Base paper achieved superior results compared to other approaches, primarily due to the use of advanced image processing techniques applied before feeding the data into the model. These preprocessing steps likely enhanced feature clarity and contributed to improved model performance. By leveraging more advanced image enhancement techniques, we were able to achieve even better results than those reported in the base paper, demonstrating the effectiveness of our improved preprocessing pipeline.

Even before applying CLAHE, our model outperformed the base paper, achieving an accuracy of 91.13%, a result attributed to the use of a more advanced model and optimized hyperparameters. However, after incorporating CLAHE into our preprocessing pipeline, the accuracy significantly improved, reaching 96.06. As observed from the confusion matrix of EfficientNetB6 before applying CLAHE, the model correctly classified 633 out of 639 Healthy images, 192 out of 223 Moderate cases, and 28 out of 52 Severe cases. These results indicate strong performance, particularly for the Healthy class, even before contrast enhancement was introduced. After applying CLAHE we got more balanced results among all the classes where we have now 625 correctly predicted out of 639 for healthy, 212 out of 223 for moderate and 42 out of 52 for severe showing that after applying CHALE model was able to learn all the complex features in images and identify all the classes correctly because of improved image contrast and lesser noise. Moreover, the model's balanced performance across all the classes and subtle difference between training and validation loss at the end of training shows the model did not overfit. Lower values of evaluation matrices for severe class is due the fact that samples for severe class in original dataset were very low compared to other classes and we performed huge amount of augmentation for severe class, the values of precision, recall and f1-score for severe class can be further enhanced by using more advanced data augmentation techniques such as elastic transformation, color jittering and random cropping. Techniques like SMOTE (Synthetic Minority Over Sampling) for generating synthetic samples can be used along with data augmentation to achieve more balanced data. EfficientNetB6 demonstrates robust performance for classification of Rheumatoid Arthritis severity and can be used as an asset for medical image classification, increasing the patient's outcomes and advancing the field of medical diagnosis.

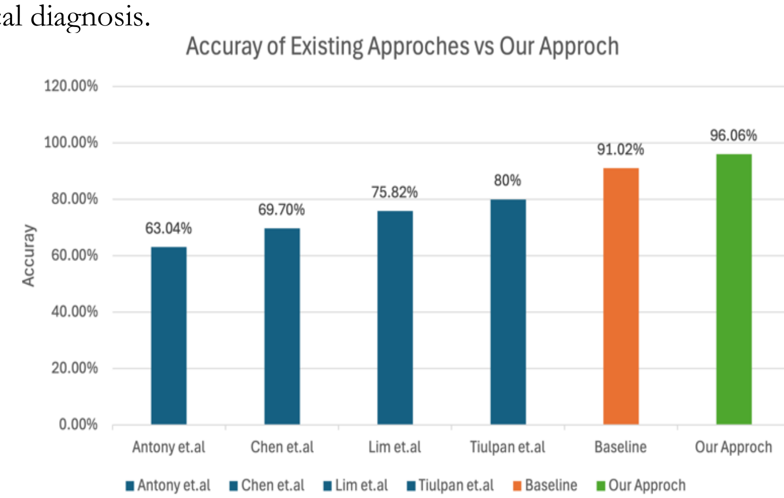

Figure 14. Comparative Analysis.

Conclusion:

In this work, we utilized a transfer learning approach using the EfficientNet family that was already trained on ImageNet data with higher results compared to other Models for Image classification. We used the X-ray images as input and first performed all the data preprocessing, including data balancing using Augmentation, and then enhancing the images using Contrast Limited Adaptive Histogram Equalization, which caused significant improvements in the results. We performed intensive hyper-parameter optimization and added robust regularizations. We performed experiments on all the EfficientNet Family models along with other state-of-the-art models for Rheumatoid Arthritis classification. EfficientNetB6 demonstrated robust performance in early detection and diagnosis of Rheumatoid Arthritis into three distinct classes: Healthy, Moderate, and Severe, with a total accuracy of 96.06% surpassing other approaches for classification of Rheumatoid Arthritis.

Higher values of all the evaluation matrices have shown that it can be used as a reliable tool in hospitals for early detection of RA by rheumatologists and radiologists. Healthcare facilities, especially in remote and rural areas, lack specialists like radiologists and rheumatologists. In these settings, the proposed system can be very helpful. Even a general physician using our system can easily diagnose with the help of just X-ray whether the patient has RA, and if yes, then what is the severity and type of intervention needed. For expert radiologists and rheumatologists, it can aid in their decision-making process [25] for accurate diagnosis. Lower value of evaluation matrices for severe class, especially f1-score, shows more room for improvement using more advanced data augmentation techniques and generating synthetic data using SMOTE, so the proposed work can be extended further to incorporate this aspect.

Acknowledgements: We would like to express our sincere gratitude and deep appreciation to Osteoarthritis Initiative (OAI) for providing the valuable source of data containing knee X-ray images, which we used to build our research. We appreciate their efforts to make this data public for researchers and scientists to make invaluable contributions in the field of Medical Imaging.

Author's Contributions: All authors have provided equal contributions to this work.

Conflict of Interest: No author of this work has any conflict of interest that could have any potential conflict with this work.

References:

[1] B. Heidari, "Rheumatoid Arthritis: Early diagnosis and treatment outcomes," *Casp. J. Intern. Med.*, vol. 2, no. 1, p. 161, Dec. 2011, Accessed: Jul. 09, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC3766928/

[2] D. A. F. J Michelle Kahlenberg, "Advances in the Medical Treatment of Rheumatoid Arthritis," *Hand Clin*, vol. 27, no. 1, pp. 11–20, 2011, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC3135413/

[3] G. S. Firestein, "Evolving concepts of rheumatoid arthritis," *Nature*, vol. 423, no. 6937, pp. 356–361, May 2003, doi: 10.1038/NATURE01661,.

[4] J. Gao, Q. Jiang, B. Zhou, and D. Chen, "Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview," *Math. Biosci. Eng.*, vol. 16, no. 6, pp. 6536–6561, 2019, doi: 10.3934/MBE.2019326,.

[5] T. B. Hongming Chen, Ola Engkvist, Yinhai Wang, Marcus Olivecrona, "The rise of deep learning in drug discovery," *Drug Discov. Today*, vol. 23, no. 6, pp. 1241–1250, 2018, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1359644617303598

[6] I. B. M. Josef S Smolen, Daniel Aletaha, "Rheumatoid arthritis," *Lancet*, vol. 388, no. 10055, pp. 2023–2038, 2016, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27156434/

[7] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017, doi: https://doi.org/10.1016/j.media.2017.07.005.

[8] D. L. Orange DE, Agius P, DiCarlo EF, Robine N, Geiger H, Szymonifka J, McNamara

M, Cummings R, Andersen KM, Mirza S, Figgie M, Ivashkiv LB, Pernis AB, Jiang CS, Frank MO, Darnell RB, Lingampali N, Robinson WH, Gravallese E; Accelerating Medicines Partnership i, "Identification of Three Rheumatoid Arthritis Disease Subtypes by Machine Learning Integration of Synovial Histologic Features and RNA Sequencing Data," *Arthritis Rheumatol*, vol. 70, no. 5, pp. 690–701, 2018, doi: 10.1002/art.40428.

[9] L. Shamir *et al.*, "Knee X-ray image analysis method for automated detection of osteoarthritis," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 2, pp. 407–415, Feb. 2009, doi: 10.1109/TBME.2008.2006025,.

[10] D. S. Teoh YX, Lai KW, Usman J, Goh SL, Mohafez H, Hasikin K, Qian P, Jiang Y, Zhang Y, "Discovering Knee Osteoarthritis Imaging Features for Diagnosis and Prognosis: Review of Manual Imaging Grading and Machine Learning Approaches," *J Heal. Eng*, 2022, doi: 10.1155/2022/4138666.

[11] N. T. A. Alexopoulos, J. Hirvasniemi, S. Klein, C Donkervoort, E.H.G. Oei, "Early detection of knee osteoarthritis using deep learning on knee magnetic resonance images," *Osteoarthr. Imaging*, vol. 3, no. 1, p. 100112, 2023, doi: https://doi.org/10.1016/j.ostima.2023.100112.

[12] J. Antony, K. McGuinness, N. E. O'Connor, and K. Moran, "Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks," *Proc. - Int. Conf. Pattern Recognit.*, vol. 0, pp. 1195–1200, Jan. 2016, doi: 10.1109/ICPR.2016.7899799.

[13] S. Murakami, K. Hatano, J. Tan, H. Kim, and T. Aoki, "Automatic identification of bone erosions in rheumatoid arthritis from hand radiographs based on deep convolutional neural network," *Multimed. Tools Appl.*, vol. 77, no. 9, pp. 10921–10937, May 2018, doi: 10.1007/S11042-017-5449-4/METRICS.

[14] L. C. Wang HJ, Su CP, Lai CC, Chen WR, Chen C, Ho LY, Chu WC, "Deep Learning-Based Computer-Aided Diagnosis of Rheumatoid Arthritis with Hand X-ray Images Conforming to Modified Total Sharp/van der Heijde Score," *Biomedicines*, vol. 10, no. 6, p. 1355, 2022, doi: 10.3390/biomedicines10061355.

[15] A. S. Jose L. Salmeron, Samira Abbasgholizadeh Rahimi, Amir Mohammad Navali, "Medical diagnosis of Rheumatoid Arthritis using data driven PSO–FCM with scarce datasets," *Neurocomputing*, vol. 232, pp. 104–112, 2017, doi: https://doi.org/10.1016/j.neucom.2016.09.113.

[16] O. I. Dziekan T, Weissbach C, Voigt J, Ebert B, Macdonald R, Bahner ML, Mahler M, Schirner M, Berliner M, Berliner B, Osel J, "Detection of rheumatoid arthritis by evaluation of normalized variances of fluorescence time correlation functions," *J Biomed Opt*, vol. 16, no. 7, p. 076015, 2011, doi: 10.1117/1.3599958.

[17] R. M. Bernd Ebert, Thomas Dziekan, Carmen Weissbach, Marianne Mahler, Michael Schirner, Birgitt Berliner, Daniel Bauer, Jan Voigt, Michael Berliner, Malte L. Bahner, "Detection of rheumatoid arthritis in humans by fluorescence imaging," *Proc. SPIE*, vol. 7555, 2010, doi: https://doi.org/10.1117/12.840913.

[18] A. Frize, Monique ; Adéa, Cynthia ; Payeur, Pierre ; Di Primio, Gina ; Karsh, Jacob ; Ogungbemile, "Detection of rheumatoid arthritis using infrared imaging," *Proc. SPIE*, vol. 7962, 2011, doi: 10.1117/12.874552.

[19] A. Das M, Ganesh Kumar; Goswami, "Automatic Classification of the Severity of Knee Osteoarthritis Using Enhanced Image Sharpening and CNN," *Appl. Sci.*, vol. 13, no. 3, p. 1658, 2023, doi: 10.3390/app13031658.

[20] Y. H. and F. W. T. Yang, H. Zhu, X. Gao, Y. Zhang, "Grading of Metacarpophalangeal Rheumatoid Arthritis on Ultrasound Images Using Machine Learning Algorithms," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2982027.

[21] R. T. Wahyuningrum, L. Anifah, I. K. Eddy Purnama, and M. Hery Purnomo, "A New

Approach to Classify Knee Osteoarthritis Severity from Radiographic Images based on CNN-LSTM Method," *2019 IEEE 10th Int. Conf. Aware. Sci. Technol. iCAST 2019 - Proc.*, Oct. 2019, doi: 10.1109/ICAWST.2019.8923284.

[22] M.-P. J. Bonakdari H, Jamshidi A, Pelletier JP, Abram F, Tardif G, "A warning machine learning algorithm for early knee osteoarthritis structural progressor patient screening," *Ther Adv Musculoskelet Dis*, 2021, doi: 10.1177/1759720X21993254.

[23] Abdullah, S.S.; Rajasekaran, M.P. Automatic detection and classification of knee osteoarthritis using deep learning approach. Radiol. Med. 2022, 127, 398–406.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, May 2022. doi 10.1038/nature14539

[25] Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (2019). Explainable AI: interpreting, explaining, and visualizing deep learning. Springer.