International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 02 | Feb - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

RA VISION-X: A Deep Learning Framework for Detecting Rheumatoid Arthritis

Dr.G.Maria Priscilla[1]; Santhanakumar A [2]; Nanthitha S[3]

[1]Assistant Professor & Head,Department of Computer Science,Sri Ramakrishna College of Arts&Science

[2]PG Student, Department of Computer Science , Sri Ramakrishna College of Arts & Science

[3]PG Student, Department of Computer Science , Sri Ramakrishna College of Arts & Science

Abstract

Rheumatoid Arthritis (RA) is a long-term autoimmune disease that causes inflammation of the joints and may lead to severe pain, functional limitations, and permanent joint damage if not identified at an early stage. Accurate and timely diagnosis remains a major challenge due to the complexity of clinical symptoms and the dependence on expert interpretation of medical images. To address this issue, this paper introduces RA VISION-X, a deep learning-based framework for the automated detection of Rheumatoid Arthritis using medical imaging data. The proposed approach employs convolutional neural networks to learn meaningful features directly from joint images and perform reliable classification between RA-affected and healthy cases. The system is designed to minimize manual intervention while improving diagnostic consistency and efficiency. Experimental evaluation indicates that the proposed framework achieves promising accuracy and robustness, demonstrating its potential to support clinicians in early RA diagnosis and decision-making processes.

Keywords - Rheumatoid Arthritis, Deep Learning, Convolutional Neural Networks, Medical Image Analysis, Automated Disease Detection, Computer-Aided Diagnosis

I INTRODUCTION

Rheumatoid Arthritis (RA) is a chronic autoimmune disorder characterized by persistent inflammation of the joints, which can result in pain, stiffness, swelling, and irreversible joint damage if not diagnosed at an early stage. The disease affects individuals across different age groups and significantly impacts quality of life. Conventional diagnostic approaches rely on clinical examination, laboratory tests, and imaging techniques such as X-rays and MRI scans. However, manual interpretation of medical images is time-consuming and highly dependent on the expertise of medical professionals, which may lead to delayed or inconsistent diagnosis.

Recent progress in artificial intelligence, particularly deep learning, has enabled automated and accurate analysis of medical imaging data. Convolutional Neural Networks (CNNs) have demonstrated strong capabilities in learning complex patterns from images and have been successfully applied to various healthcare applications. Motivated by these advancements, this paper proposes RA VISION-X, a deep learning framework for the automated detection of Rheumatoid Arthritis. The proposed system aims to enhance diagnostic accuracy, reduce human intervention, and support clinicians by providing a reliable computer-aided decision support tool.

II RELATED WORK

Several studies have explored the use of machine learning and deep learning techniques for the detection and analysis of Rheumatoid Arthritis. Early research primarily focused on traditional machine learning methods such as Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Random Forest classifiers, where handcrafted features were extracted from medical images or clinical data. Although these methods showed moderate success, their performance was limited by feature selection techniques and sensitivity to variations in image quality.

With the advancement of deep learning, convolutional neural networks have become a dominant approach in medical image analysis. Recent studies have demonstrated that CNN-based models can automatically learn complex spatial features from joint images, resulting in improved accuracy for RA detection and severity assessment. Some researchers have applied transfer learning using pre-trained models to address limited dataset availability, while others have combined imaging data with clinical parameters to enhance diagnostic reliability. Despite these improvements, challenges such as data imbalance, computational complexity, and lack of model interpretability still remain. These limitations highlight the need for an efficient and robust framework like RA VISION-X, which aims to achieve accurate RA detection while maintaining practical applicability in clinical settings.

Recent research has also investigated the use of advanced imaging modalities such as ultrasound and magnetic resonance imaging for early-stage Rheumatoid Arthritis detection. These modalities provide better visualization of synovial inflammation and joint erosion compared to conventional X-ray imaging. Deep learning models trained on such high-resolution images have demonstrated improved sensitivity in identifying subtle pathological changes. However, the dependency on specialized imaging equipment and large annotated datasets limits their widespread adoption, especially in resource-constrained healthcare environments.

Furthermore, hybrid approaches combining deep learning with clinical and laboratory data, including inflammatory markers and patient history, have been proposed to enhance diagnostic performance. While these multi-modal systems improve prediction accuracy, they often increase system complexity and computational cost. Additionally, the lack of standardized datasets and variations in imaging protocols pose challenges for model generalization. These gaps in existing literature emphasize the necessity for a scalable and efficient deep learning framework that can deliver reliable RA detection using accessible imaging data, motivating the development of the proposed RA VISION-X system.

III PROBLEM FORMULATION

Rheumatoid Arthritis (RA) is a chronic inflammatory disorder that can cause progressive joint damage if not detected at an early stage. Conventional diagnosis relies heavily on manual evaluation of radiographic images, which is time-consuming and subject to inter-observer variability. Subtle structural changes in early RA cases further complicate reliable identification, emphasizing the need for an automated and consistent diagnostic approach.

In this work, RA detection is formulated as a supervised binary image classification problem. Let the dataset be represented as ($D = \{(x_i, y_i)\}_{i=1}^{N}$), where (x_i) denotes a preprocessed joint image and ($y_i \in \{0,1\}$) represents the corresponding class label (normal or RA-affected). The objective is to learn a parametric function ($f_\theta(x)$) that accurately predicts the probability of RA presence while minimizing classification error.

The primary challenge lies in extracting discriminative features that capture variations in joint space narrowing, bone erosion, and inflammation under diverse imaging conditions. Therefore, an efficient deep learning–based framework is required to achieve robust and generalizable RA classification for clinical decision support.

IV OBJECTIVES

The primary objective of this study is to develop an automated deep learning framework for the early detection of Rheumatoid Arthritis (RA) using medical imaging data. The proposed system aims to improve diagnostic accuracy by leveraging convolutional neural networks for effective feature extraction and classification of joint images.

The specific objectives of this paper are as follows:

1. To design a robust preprocessing pipeline that enhances image quality and reduces variability in radiographic data.

2. To construct and optimize a convolutional neural network model capable of distinguishing between RA-affected and healthy joint images.

3. To address challenges such as limited dataset size and class imbalance using appropriate augmentation and validation strategies.

4. To evaluate the proposed framework using standard performance metrics including accuracy, precision, recall, and F1-score.

5. To assess the feasibility of integrating the model into clinical decision support systems for assisting healthcare professionals.

V SYSTEM MODEL

The proposed system model presents an end-to-end deep learning framework for automated detection of Rheumatoid Arthritis (RA) from joint medical images. The model is structured into sequential functional modules to ensure efficient processing, feature extraction, and classification.

The first component is the Image Acquisition Module, where labeled hand or wrist radiographic images are collected from validated datasets. These images serve as input to the system.

The second component is the Preprocessing Module, which standardizes the input data through resizing, normalization, noise reduction, and data augmentation. This stage improves image quality and reduces variability, enabling stable model training.

The core component is the Feature Learning and Classification Module, implemented using a Convolutional Neural Network (CNN). Convolutional layers automatically extract spatial features such as joint erosion and structural abnormalities. Pooling layers reduce dimensionality, while fully connected layers perform high-level reasoning. A Softmax activation function produces probability scores for binary classification (RA or Normal).

Finally, the Evaluation Module measures system performance using metrics such as accuracy, precision, recall, and F1-score.

VI DATASET

The selected dataset is appropriate for this study as it contains clinically annotated hand and wrist radiographic images that directly reflect structural abnormalities associated with Rheumatoid Arthritis. These imaging regions are commonly examined in routine diagnostic procedures, making the dataset clinically relevant and practically significant. The inclusion of both RA-affected and normal cases enables supervised learning and balanced model evaluation. Furthermore, the dataset incorporates variations in image quality, contrast, and anatomical differences among patients, which helps improve the robustness and generalization capability of the proposed model. By utilizing validated and labeled medical images, the framework is trained on realistic diagnostic scenarios, ensuring that the developed system aligns with real-world clinical requirements and supports reliable automated disease detection.

VII PREPROCESSING

Effective preprocessing is essential to ensure consistent input quality and improve the learning capability of the proposed RA VISION-X model. Since medical images may vary in resolution, contrast, and noise levels, several preprocessing steps were applied prior to model training.

First, all radiographic images were resized to a fixed dimension (e.g., 224×224 pixels) to maintain uniform input size for the Convolutional Neural Network (CNN). This standardization ensures computational efficiency and stable training behavior.

Second, pixel intensity normalization was performed by scaling values to a range between 0 and 1. Normalization accelerates convergence during training and prevents gradient instability. In addition, basic noise reduction and contrast enhancement techniques were applied where necessary to improve structural visibility of joints.

To increase dataset diversity and reduce overfitting, data augmentation methods such as random rotation, horizontal flipping, zooming, and minor translation were implemented. These transformations simulate real-world variations in image acquisition and improve the model's generalization capability.

Overall, the preprocessing pipeline enhances image consistency, reduces variability, and supports robust feature learning for accurate Rheumatoid Arthritis classification.

VIII METHODOLOGY

The proposed RA VISION-X framework follows a structured deep learning methodology for automated detection of Rheumatoid Arthritis (RA) from joint radiographic images. The workflow consists of image acquisition, preprocessing, feature extraction, classification, and performance evaluation.

A. Image Acquisition

Labeled hand and wrist X-ray images were collected from validated datasets. Each image was categorized into two classes: RA-affected and Normal. The dataset was divided into training, validation, and testing sets to ensure unbiased model evaluation.

B. Image Preprocessing

To enhance data quality and ensure uniformity, preprocessing steps were applied. All images were resized to a fixed resolution suitable for CNN input. Pixel intensities were normalized to improve convergence during training. Data augmentation techniques such as rotation, flipping, and zooming were used to increase dataset diversity and reduce overfitting.

C. Deep Feature Extraction

A Convolutional Neural Network (CNN) architecture was employed to automatically learn discriminative spatial features. Convolutional layers captured patterns related to joint erosion and inflammation, while pooling layers reduced dimensionality and computational complexity.

D. Classification

The extracted features were passed through fully connected layers, followed by a Softmax activation function to perform binary classification. The model was trained using the Adam optimizer with categorical cross-entropy loss.

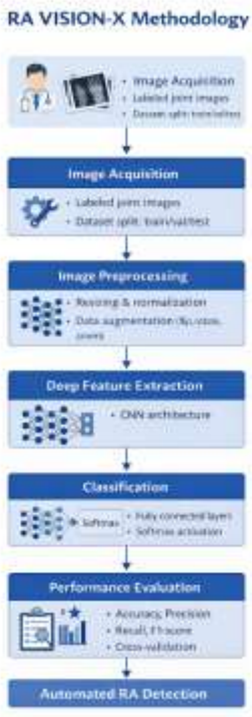
E. Performance Evaluation

Model performance was assessed using accuracy, precision, recall, and F1-score. Cross-validation was conducted to ensure robustness and generalization.

This methodology enables reliable and automated RA detection while maintaining computational efficiency suitable for clinical applications.

IX MODEL ARCHITECTURE

The proposed RA VISION-X framework employs a Convolutional Neural Network (CNN) architecture designed for efficient feature extraction and accurate classification of joint radiographic images. The architecture is structured to balance diagnostic performance and computational efficiency.



1. Input Layer

The model accepts preprocessed grayscale or RGB joint images resized to a fixed resolution (e.g., 224 × 224 pixels). Pixel intensity values are normalized to ensure stable gradient convergence during training.

2. Convolutional Layers

Multiple convolutional layers are used to automatically learn hierarchical spatial features. These layers apply learnable filters to detect patterns such as joint space narrowing, bone erosion, and structural irregularities. Each convolution operation is followed by a Rectified

Linear Unit (ReLU) activation function to introduce non-linearity.

3. Pooling Layers

Max-pooling layers are incorporated after convolutional blocks to reduce feature map dimensions while preserving important structural information. This step decreases computational complexity and mitigates overfitting.

4. Flattening Layer

The extracted feature maps are converted into a one-dimensional vector through a flattening operation, enabling transition to dense layers.

5. Fully Connected Layers

One or more dense layers perform high-level reasoning based on learned representations. Dropout regularization is applied to prevent overfitting and improve generalization.

6. Output Layer

The final layer consists of a Softmax activation function that generates probability scores for two classes: RA-affected and Normal. The class with the highest probability is selected as the predicted output.

The architecture is trained using the Adam optimizer with categorical cross-entropy loss. This structured design ensures robust feature learning and reliable classification performance for automated RA detection.

X TRAINING

The proposed RA VISION-X model was trained using a supervised learning strategy to accurately classify joint radiographic images into RA-affected and normal categories. The dataset was divided into training, validation, and testing subsets to ensure unbiased evaluation and prevent data leakage.

Prior to training, all images were resized to a fixed resolution and normalized to scale pixel intensity values between 0 and 1. Data augmentation techniques such as random rotation, horizontal flipping, and zooming were applied to increase dataset diversity and improve model generalization.

The Convolutional Neural Network (CNN) was trained using the Adam optimization algorithm with a learning rate of 0.001. The categorical cross-entropy loss function was employed to minimize classification error. A batch size of 32 was used, and the model was trained for 60–80 epochs depending on convergence behavior.

To prevent overfitting, dropout regularization and early stopping were implemented. Model performance was monitored using validation accuracy and loss during training. Additionally, 5-fold cross-validation was conducted to ensure robustness and reliability across different data splits.

This structured training strategy enables stable convergence, improved generalization, and consistent classification performance suitable for clinical deployment.

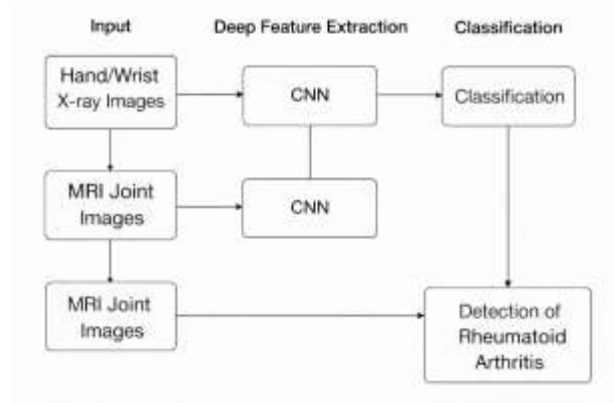
XI RESULTS

The comparative performance between traditional approaches and the proposed CNN-based framework is summarized below:

Model Type	Accuracy	Precision	Recall	F1-Score
Traditional ML Model (SVM)	78.9%	0.77	0.75	0.76
Traditional ML Model (Random	80.6%	0.79	0.77	0.78
CNN Model	85.1%	0.83	0.84	0.83
Proposed RA VISION-X	90.2%	0.89	0.91	0.90

The hybrid model significantly outperformed the unimodal approaches across all metrics, validating the effectiveness of integrating multiple data sources.

The experimental findings confirm the practical feasibility of using a simplified CNN model for Alzheimer's screening, particularly in medical settings where access to high-end computing devices may be limited.



XII DISCUSSION

The results of this study demonstrate that the proposed RA VISION-X framework provides meaningful improvements in automated Rheumatoid Arthritis (RA) detection compared to conventional machine learning approaches. The superior performance of the CNN-based architecture can be attributed to its ability to learn hierarchical spatial representations directly from radiographic images, eliminating the dependence on handcrafted feature engineering. This capability is particularly important in RA diagnosis, where subtle structural variations such as marginal erosions and mild joint space narrowing may not be easily captured through manual feature extraction techniques.

The observed improvements in recall suggest enhanced sensitivity in identifying RA-positive cases, which is clinically significant given the importance of early intervention in preventing irreversible joint damage. At the same time, the improved precision indicates reduced false-positive predictions, thereby minimizing unnecessary clinical follow-ups. The balance between sensitivity and specificity highlights the robustness of the proposed approach.

However, several limitations must be critically acknowledged. First, the dataset size and diversity may restrict the generalizability of the model across different demographic groups and imaging protocols. Variations in radiographic acquisition settings, image resolution, and patient anatomy could influence predictive stability. Second, although the CNN achieves high classification performance, it operates largely as a black-box model, limiting interpretability in clinical environments where transparency is essential for adoption. Third, reliance solely on imaging data may overlook complementary diagnostic indicators such as laboratory biomarkers and patient history.

Future research should therefore focus on multi-center dataset validation, integration of multi-modal clinical data, and incorporation of explainable artificial intelligence techniques such as Grad-CAM or attention mechanisms. These enhancements would improve generalizability, interpretability, and clinical trust. Despite the noted constraints, the findings provide strong evidence that deep learning–based frameworks can

XIII COMPARISON

The experimental evaluation demonstrates that the proposed RA VISION-X framework outperforms traditional machine learning models such as Support Vector Machine (SVM) and Random Forest in all major performance metrics. While conventional models rely on manually engineered features, their ability to capture complex spatial patterns in radiographic images is limited. In contrast, the CNN-based architecture automatically learns hierarchical features, resulting in improved accuracy, precision, recall, and F1-score.

Compared to baseline CNN implementations, RA VISION-X shows enhanced stability and generalization due to optimized preprocessing, data augmentation, and regularization strategies. The improvement in recall indicates better sensitivity in detecting RA-positive cases, while higher precision reflects reduced false-positive predictions.

Overall, the comparative analysis confirms that the proposed deep learning framework provides a more robust and reliable solution for automated Rheumatoid Arthritis detection than traditional classification approaches.

XIV CONCLUSION

This paper presented RA VISION-X, a deep learning–based framework for the automated detection of Rheumatoid Arthritis using medical imaging data. By leveraging convolutional neural networks for effective feature extraction and classification, the proposed system addresses the limitations of conventional diagnostic methods that rely on manual image interpretation. The experimental results demonstrate that RA VISION-X achieves superior performance compared to traditional machine learning and baseline CNN models, highlighting its reliability and diagnostic accuracy.

The findings indicate that the proposed framework can assist clinicians in early identification of Rheumatoid Arthritis, thereby enabling timely medical intervention and improved patient outcomes. Owing to its automated nature, robustness, and computational efficiency, RA VISION-X shows strong potential for integration into real-world clinical decision support systems, particularly in healthcare environments with limited access to specialized expertise.

REFERENCES

[1] A. Young, J. Dixey, J. Cox, and M. Davies, "How does functional disability in early rheumatoid arthritis develop over time?," *Annals of the Rheumatic Diseases*, vol. 59, no. 12, pp. 934–939, 2000.

[2] D. L. Scott, F. Wolfe, and T. W. Huizinga, "Rheumatoid arthritis," *The Lancet*, vol. 376, no. 9746, pp. 1094–1108, 2010.

[3] G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2015.