Journal of Universal Computer Science, vol. 32, no. 3 (2026), 373-404
submitted: 26/6/2024, accepted: 18/2/2025, appeared: 28/3/2026

Enhancing Rheumatoid Arthritis Diagnosis: Combining Case-Based Reasoning on EHR Data with Deep Learning on Medical Images

Moulay Youssef Ichahane
(LTI/SI /CED Technology and engineering, ENSA El Jadida/El Jadida, Morocco
https://orcid.org/0000-0001-7382-1019, y.ichahane@gmail.com)

Noureddine Assad
(LTI/SI /CED Technology and engineering, ENSA El Jadida/El Jadida, Morocco
https://orcid.org/0000-0002-1662-0066, assad.noureddine@gmail.com)

Hassan Ouahmane
(LTI/SI /CED Technology and engineering, ENSA El Jadida/El Jadida, Morocco
https://orcid.org/0000-0001-8723-6590, hassan.ouahmane@yahoo.fr)

Abstract: The diagnosis of rheumatoid arthritis (RA) diagnosis demands precise detection methods due to its complex symptomatology. This study presents a novel hybrid diagnostic framework that is the first to integrates Case-Based Reasoning (CBR) with deep learning and introduce three key innovations: (i) a dual-pathway architecture that combine electronic health records with imaging analysis, (ii) an Enhanced Clustering-Based K-nearest neighbors (ECB KNN) model for optimal feature selection, and (iii) a dynamic K-means clustering approach for handling class imbalance. We evaluated our framework using two comprehensive datasets: MIMIC-IV-Hosp, containing clinical data and MIMIC-CXR containing 377,110 chest X-ray images. The model employs a VGG16-based CNN for radiological feature extraction, with a particular focus on pulmonary manifestations, which is combined with our ECB KNN algorithm for patient-specific clinical data analysis. Using five-fold cross-validation, our framework is shown to achieve superior performance metrics (precision: 0.90-0.95, recall: 0.89-0.93, F1-score: 0.91) compared to baseline methods (traditional CNN: precision 0.82, recall 0.79; standard CBR: precision 0.85, recall 0.83). This significant improvement in diagnostic accuracy demonstrates the potential of our framework in terms of enhancing early RA detection and clinical decision support. The architecture of the model architecture is designed to allow for extensibility to other rheumatic conditions, thereby offering a comprehensive solution for multi-disease diagnosis in rheumatology.

Keywords: Rheumatoid Arthritis, Diagnosis, Case-Based Reasoning, HER, Deep Learning, Healthcare, Precision Medicine, Disease Detection, Computer-Aided Diagnosis
Categories: I.5.3, J.3, K.6.2, C.4, K.6.2, K.6.4, I.2.7, J.3
DOI: 10.3897/jucs.130529

1 Introduction

Rheumatoid Arthritis (RA) is a significant global health challenge that affect millions worldwide and gives rise to complex diagnostic needs due to its systemic autoimmune nature and symptoms that overlap with those of other diseases [World Health Organization, 2024]. Traditional diagnostic methods, which are primarily based on clinical assessment, laboratory results, and imaging, are often insufficient for early detection and precise diagnosis, particularly in complex cases with secondary manifestations such as interstitial lung disease (ILD) or pleural involvement [Cala et al., 2022]. The World Health Organization has emphasised the urgency of addressing RA by ranking it among the top ten causes of disability, highlighting the need for effective diagnosis and treatment strategies [World Health Organization, 2024]. Approximately 18 million people were living with RA in 2019 of which the majority were women and more than half were over 55 years old, and many of these individuals experience levels of severity that would benefit from rehabilitation interventions. Despite advances in biological and synthetic treatments, a significant reduction in progression of the disease is achieved only in a small subset of patients.

The evolution of medical science has ushered in evidence-based medicine, which has fundamentally transformed the diagnostic landscape. Precise and timely diagnosis of RA, which was previously only possible through subjective symptom assessment, now relies on methodologies grounded in empirical evidence, significantly impacting patient outcomes and quality of life [Marques et al., 2024]. RA often shares clinical features with related autoimmune diseases such as systemic lupus erythematosus and psoriatic arthritis, making differential diagnosis complicated [Cala et al., 2022]. This complexity calls for a more refined approach that combines unique clinical symptoms with biological markers, and radiological evidence to improve diagnostic precision.

The introduction of artificial intelligence (AI) into healthcare has given rise to opportunities to enhance diagnostic accuracy and speed, which are crucial for the effective management of RA and improved patient outcomes [Giovanna et al., 2024]. In recent years, deep learning (DL) and case-based reasoning (CBR) have shown promise in medical diagnostics. DL has proven effective in feature extraction from medical images, while CBR enables personalised decision support by leveraging historical cases [Momtazmanesh et al., 2022]. This study builds on these advancements by introducing a novel, integrated framework for RA diagnosis. By combining the image-based pattern recognition capabilities of DL with the knowledge-driven reasoning of CBR, the proposed framework aims to address the diagnostic challenges associated with RA, with a specific focus on the integration of chest X-ray analysis with electronic health record (EHR) data. However, our approach is not limited to conventional methods, and we aim to reduce the list of potential diagnoses based on individual patient symptoms and biological data [Momtazmanesh et al., 2022]. This personalised approach significantly refines the diagnostic process. Furthermore, we expect that the fusion of CBR with a convolutional neural network (CNN) for radiological images will not only enhances the accuracy of disease detection but also provide insight into the stage of progression of the disease.

A literature survey reveals several critical implications and gaps in the current approaches to RA diagnosis using CBR and DL. One major gap that can be identified is the integration and methodology conflicts that are inherent in the application of CBR, as effective utilization requires careful consideration of the relationships within multi-level diagnostic systems, which can lead to methodological inconsistencies[Scully, 2006].In addition, although a combination of CBR with neural networks has shown potential in regard to predicting early arthritis, the accuracy and confidence levels will need further improvement before this approach becomes reliable for clinical use. The lack of specific DL applications in RA diagnosis is another gap, as previous work has largely focused on other areas of medical diagnostics, such as COVID-19 epidemiological surveys [Lin et al., 2024].

In summary, our approach addresses the methodological conflicts and accuracy issues associated with CBR and explores the underutilised potential of DL in RA diagnosis. By combining these techniques and incorporating a diverse range of clinical data, we aim to create an AI-driven tool that is not only accurate but also interpretable and aligned with clinical needs, thereby advancing the practical application of machine learning in RA diagnostics.

Motivated by the recent introduction of AI into the field of healthcare, our work advances medical diagnostics as follows:

(i) We introduce a novel integration methodology that dynamically fuses CBR and CNN pathways, with their contributions being adapted based on data quality and diagnostic confidence.

(ii) We implement an adaptive feature importance scoring method, guided by the Maximal Information Coefficient (MIC), which automatically adjusts to evolving disease patterns.

(iii) We introduce a bi-directional feedback loop in which CNN-derived features enhance the CBR similarity assessments, while insights from the CBR system are used to refine the CNN feature extraction process. The CNN is retrained to emphasize these critical features, with a particular focus on lung areas during image analysis, based on tuning of the attention mechanisms within the CNN for a more intensive focus,

(iv) Our dynamic clustering-enhanced KNN evolves with each new case, effectively addressing the issue of class imbalance through integrated synthetic sampling and merging learned and expert-defined features for more nuanced similarity metrics.

(v) The enhanced VGG16 architecture includes custom attention mechanisms that are specifically tailored to RA manifestations, domain-specific skip connections, and a severity-weighted loss function for more targeted learning.

(vi) In a clinical context, we present the first successful integration of DL-based radiological analysis with CBR for RA diagnosis,

(vii) We propose an adaptive feature compensation strategy to handle the problem of missing data effectively,

(viii) We demonstrate a real-time explainable AI system that is validated in a clinical setting.

These contributions collectively advance both the theoretical and practical dimensions of RA diagnostics, providing a comprehensive, precise, and interpretable solution.

The article begins with an Introduction presenting the study's motivation, background, and contributions, followed by a Related Work section reviewing evidence-based medicine for diagnosis, current RA diagnostic methods, the complexity of RA, and recent DL applications. The Materials and Methods section detail the datasets, preprocessing strategies, data validation, and the efficient treatment identification framework. The Proposed Framework describes the CNN architecture, the CBR technique, and the clustering-enhanced KNN with CNN integration. The Experimental Results and Discussions present implementation details, the ILD detection methodology, and a comparative analysis of results. Finally, the Conclusion summarises the main findings, discusses their clinical implications, and outlines potential directions for future research.

2 Related work

2.1 Evidence-based Medicine for Diagnosis

It is important to note that the development of a machine learning algorithm for medical diagnosis, especially for a complex condition such as RA, requires a large and diverse dataset, collaboration with medical experts, and careful consideration of ethical and regulatory considerations.

Arthritis is a term that is used for various inflammatory conditions that affect different parts of the body such as joints, bones, and muscles. It can take several forms, such as osteoarthritis (OA), RA, juvenile arthritis, psoriatic arthritis, and gouty arthritis, and can result in stiffness, pain, redness, and swelling in the joints. Untreated, RA can cause severe damage to the joints and their surrounding tissue. It can lead to heart, lung, or nervous system problems. If diagnosed in a timely manner, the symptoms and progression of the disease can be controlled via pharmacological treatment, and optimal functioning can be maintained through rehabilitation (including the use of assistive products). In cases involving severe joint damage, surgical procedures such as joint replacement, may help to restore movement or manage pain, and maintain physical function.

Additionally, the steps in the diagnostic algorithm presented here provide only a high-level overview, and the specific details of the implementation and the choice of algorithm may vary based on the complexity of the problem and the data that are available.

Diagnosis of RA in the context of evidence-based medicine typically relies on a combination of clinical assessment, laboratory tests, and imaging. Although an algorithm alone cannot replace the expertise of a qualified healthcare provider, a simplified outline of a typical diagnostic algorithm used in evidence-based medicine for RA is given in Figure 1 below.

Figure 1: Simplified diagnostic algorithm used in evidence-based medicine for RA

We emphasize that the diagnosis of RA should be made by qualified healthcare professionals, based on a consideration of the patient's clinical presentation and a comprehensive assessment of the available data. An algorithm provides a general framework for understanding the diagnostic process, but it does not replace the expertise and judgment of medical practitioners. This approach should also be based on the most up-to-date evidence and guidelines in rheumatology.

2.2 RA diagnosis:

The current evaluation of RA integrates clinical examination, patient-reported symptoms, laboratory investigations, and imaging modalities—primarily conventional radiography and magnetic resonance imaging (MRI), with computed tomography (CT) or other techniques employed in specific clinical scenarios, in accordance with evidence-based recommendations on the role of imaging in RA management developed by an international task force of rheumatology and radiology experts [Colebatch et al., 2013]. However, the Table 1 outlines the chronological progression of radiographic pulmonary manifestations frequently associated with RA. It synthesises the principal radiological features detectable on chest imaging, ranging from early to advanced stages of pulmonary involvement. This classification serves as a valuable tool for assessing disease severity and informing appropriate clinical management strategies.

Finding	**Imaging Modality**	**Clinical Relevance**
Pleural Effusion	Chest X-ray, Ultrasound	Often the earliest thoracic manifestation in RA; suggests pleuritis
Pulmonary Nodules	Chest X-ray, CT	Can appear early, even in asymptomatic patients
Caplan Syndrome (RA + Pneumoconiosis)	X-ray, CT	Seen in RA patients with occupational exposure (e.g., coal, silica)
Bronchiectasis	CT	Develops with chronic airway inflammation; infection risk
Interstitial Lung Disease (ILD)	Chest X-ray, CT	Appears in subacute or chronic RA; most frequent lung complication
Pulmonary Fibrosis	X-ray, CT	Late stage of ILD; irreversible lung remodeling
Pulmonary Hypertension	Chest X-ray, Echo	Often a late sequela of ILD or pulmonary vasculitis

Table 1: Chronological progression of radiographic pulmonary manifestations in RA

To establish a more structured and reliable diagnostic process, there is a need for computer-assisted analysis and predictive modelling. This is crucial in order to mitigate human errors and to enable early disease detection, especially in regions where the presence of medical experts is limited.

In addition, RA is a chronic autoimmune condition that disproportionately affects women, particularly during the premenopausal years [Macejová, 2011], [Rovenský and Masaryk, 2012]. The disease typically manifests earlier in women compared to men, resulting in longer exposure to inflammatory processes and potentially more severe joint damage. Research has shown that difficult-to-treat (D2T) cases of RA are more commonly observed in female patients, who generally experience more intense symptoms and faster disease progression than men [Shouval et al., 2024]. While the exact causes of this gender disparity remain complex, scientists have identified several contributing factors, including genetic predisposition and hormonal influences, particularly in premenopausal women [Rovenský and Masaryk, 2012], [Shinde et al., 2014]. These hormonal factors are of special interest to researchers due to their potential role in immune regulation and susceptibility to autoimmune diseases such as RA.

2.3 Complex nature of RA

2.3.1 Integrating Chest X-Ray in RA Diagnosis in a CBR-EHR Framework

The integration of chest X-ray data into RA diagnosis offers substantial clinical benefits, even though chest X-rays are not typically primary diagnostic tools for RA. They can play a critical complementary role by identifying secondary pulmonary manifestations, such as interstitial lung disease, pleuritis, and nodules, which are frequently associated with the progression of RA [Matuszewska et al., 2014]; [Fukae and Koike, 2009]. These respiratory complications serve as significant markers of disease severity and can dramatically influence patient outcomes. Early detection of these conditions provides a more complete understanding of disease progression and can support more comprehensive management strategies that can address the full impact of RA.

The incorporation of chest X-ray data into a CBR system combined with EHR enhances the diagnostic framework by adding depth to patient profiling, which now encompasses both joint and pulmonary health indicators. This approach allows the system to recognise patterns across cases where RA patients exhibit both primary and secondary manifestations, thereby enabling more precise predictive modeling of disease progression [Sudoł-Szopińska and Ćwikła, 2013] [Bliddal et al., 2008]. This can enhance the system's ability to detect trends that are indicative of a patient's broader disease trajectory, thus allowing for earlier intervention and potentially mitigating the risks associated with untreated or late-stage pulmonary involvement.

The proposed CBR-EHR system architecture leverages this integrated approach to continuously adapt and refine diagnostic predictions. By learning from new case presentations, it supports pattern matching and achieves greater diagnostic accuracy by referencing a more complete patient profile. This architecture enables early, personalised intervention by referencing outcomes from similar historical cases. The integration of chest X-ray data within a CBR-EHR framework represents a significant advancement in RA diagnosis and management, with a shift in focus from isolated symptom assessment to a comprehensive evaluation that considers both joint and systemic manifestations. This approach supports a more dynamic, patient-centered model of care that adapts to the complexities of RA over time [Verghese et al., 2021] [Rajesh et al., 2024].

2.3.2 Others Risk factors

An understanding of the interplay between genetic, environmental, and lifestyle factors is crucial for the early detection and prevention of RA, since Combinations of genetic predispositions with environmental triggers, such as smoking and specific dietary factors, have been shown to significantly influence the development of the disease [Deane et al., 2017], [Romão and Fonseca, 2021]. Addressing modifiable risk factors, including cessation of smoking and maintaining a healthy weight, can potentially reduce the risk of developing RA and mitigate disease progression; [Aho and Heliövaara, 2004]. Furthermore, early detection through imaging and serological tests enables timely interventions, which are associated with improved patient outcomes and may delay or prevent the beginning of RA in at-risk populations [Mankia et al., 2021]; [Arend and Firestein, 2012].

2.4 Application of deep learning techniques to RA

The application of DL techniques in the context of diagnosing RA has garnered significant attention from researchers in recent years, and numerous studies have explored the potential of these techniques in improving the accuracy and efficiency of RA diagnosis. We review a selection of key published works that have made notable contributions to this evolving field. The authors Üreten et al. [2020] used a CNN to analyse radiographic images of joint deformities in RA patients. Their approach showed promising results in the automated identification of RA-related structural damage, with a focus on hands and wrists. In Momtazmanesh et al. [2022] and Hossain et al. [2023], researchers reported a novel deep learning approach for the early detection of RA using natural language processing techniques. They developed a model that analysed electronic health records (EHRs) and clinical notes to identify linguistic patterns and textual markers associated with RA symptoms. The authors of Zhou et al. [2022] and Zhou et al. [2021] proposed a deep learning-based system for the classification of RA disease activity. By integrating multimodal data, including clinical assessments and medical images, their model accurately categorized patients into different disease activity states, aiding in treatment decisions. In Kalweit et al. [2021] and Rasmy L et al. [2022] the use of DL models was extended to predict future RA outcomes, and it was demonstrated that recurrent neural networks could forecast disease progression and response to therapy based on historical patient data, thereby contributing to personalized treatment plans. The studies of Hossain et al. [2022] and [2021] explored the role of transfer learning in RA diagnosis, and involved pre-training deep learning models on large medical imaging datasets and fine-tuning them for RA-specific tasks, resulting in improved diagnostic accuracy. The authors Jo et al. [2022] and Al-Maini et al. [2023] investigated the integration of DL with genetic data to identify genes related to RA susceptibility. By employing CNNs and recurrent networks, they successfully identified novel genetic factors associated with the risk of RA. The synergy between DL and CBR has been explored in various domains. In the context of healthcare, studies such as [Bichindaritz, 2023] have investigated the combination of DL feature extraction capabilities with CBR's knowledge-based decision support [Mehli et al., 2021].

The combination of CBR and neural networks has been explored in the context of diagnostic systems for various diseases. We will review some examples of research and studies related to the integration of CBR with NN for diagnostic systems. In the early stages of development in this field, Reategui et al. [1997] created a model to diagnose a congenital heart disease. Their system used a neural network to make hypotheses and guide the CBR module in the search for similar previous cases that supported these hypotheses. The knowledge acquired by the network was interpreted and mapped onto symbolic diagnosis descriptors, which were used to determine the credibility of the final answer and to build explanations for the reasoning of the model. In Sharaf-El-Deen et al. [2014], researchers worked on the development of expert systems for medical diagnosis, by comparing the performance of CBR and rule-based reasoning for medical diagnosis. They then compared the performance of the expert systems using a range of metrics, including accuracy, precision, recall, and F1-score. Finally, the researchers emphasised the importance of interpretability and robustness in medical diagnostic systems, which can make expert systems, including those based on CBR, ideal choices for medical applications. The study in Momtazmanesh et al. [2022] involved a review of the use of multiple models, including CNNs, to diagnose RA based on input of consisting of hand X-ray data, and emphasised the importance of combining imaging findings with clinical data and data derived from sensors for the diagnosis, monitoring, and management of RA.

Until recently, specific examples involving a combination of CBR and CNN for RA diagnosis were very rare. Firstly, the majority of studies have focused primarily on single-modality data (either EHR or medical imaging data), without fully exploring the potential of multimodal data integration, thus limiting the completeness of diagnostic models. Secondly, there has been a lack of emphasis on the practical implementation of these models in real-world clinical settings, including aspects related to interoperability with existing healthcare systems and challenges associated with user adoption. Finally, while many models have demonstrated promising results in controlled experimental setups, their generalisability and robustness across diverse patient populations and clinical environments remain underexplored.

Recent advancements have demonstrated the efficacy of DL in analysing complex patterns within medical imaging and EHR, which has provided significant improvements in the diagnosis of RA [Li and others, 2024]. In particular, studies of multimodal approaches in which imaging data are combined with clinical records have shown promise in identifying critical RA indicators, such as pulmonary and joint manifestations, which are essential for early diagnosis and intervention. For instance, CNN-based architectures have proven effective in processing and highlighting disease-specific features in radiographs and MRI scans [Panwar et al., 2020], [Stoel et al., 2024]. However, challenges such as data imbalance and complex interpretation persist, especially in the context of minority conditions such as RA, and novel techniques such as cost-sensitive learning and SMOTE have been employed to enhance model reliability [Siddiqui et al., 2020] [Siddiqui et al., 2019]. the use of these methods underscores the importance of tailored architectures that can balance predictive performance with interpretability, an essential aspect for applications in clinical decision support systems.

3 Materials and Methods

In this case study, a system is developed for RA diagnosis in patients, based on medical data combined with X-ray images. We hypothesise that there are subtle but informative signs of RA that may not be discernible by humans or via simple measurements or visual inspection by experts, and believe that DL has the potential to use these signs in order to produce enhanced diagnosis models. A diagram to illustrate the proposed framework is shown in *Figure 2*.

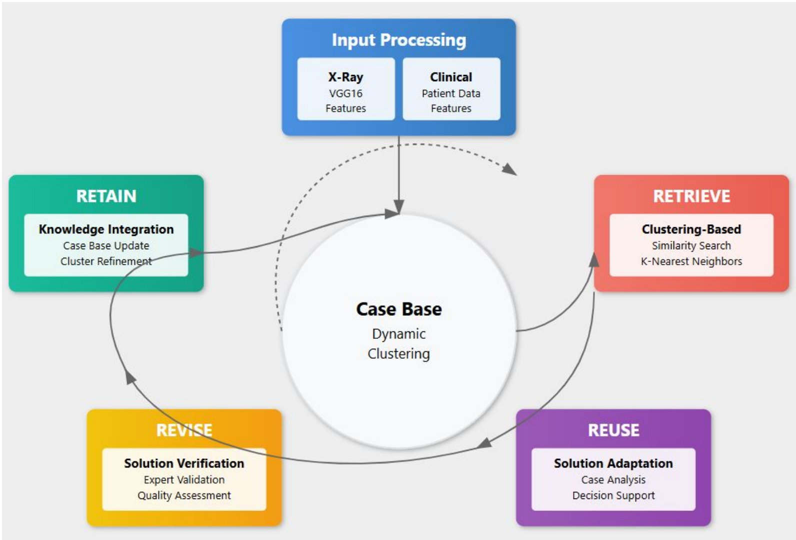

Figure 2: Conceptual architecture of the proposed case-based reasoning framework for medical diagnostics.

Our approach, illustrated in *Figure 3*, combines hospital records and chest radiographs from the MIMIC-IV-Hosp and MIMIC-CXR datasets to support the early detection of RA. After extracting relevant clinical indicators and imaging features such as ILD, nodules, and pleural effusions, the data undergo cleaning, normalization, and feature extraction using VGG16. Thereafter, the target labels are defined from clinical and radiological criteria, followed by validation to ensure consistency and reliability.

The dataset is then split into balanced training and testing sets, and enabled to two parallel modeling paths: an enhanced KNN for structured EHR data, and a CNN for imaging. The model performance is measured through 5-fold cross-validation and continuous monitoring, and fed into a comparison of different fusion strategies. The enhanced CBR method consistently outperforms the static CKC approach.

Finally, a feedback loop and continuous learning process allow the system to adapt over time, with the results consolidated into a final performance report that includes metrics, and clinical validation.

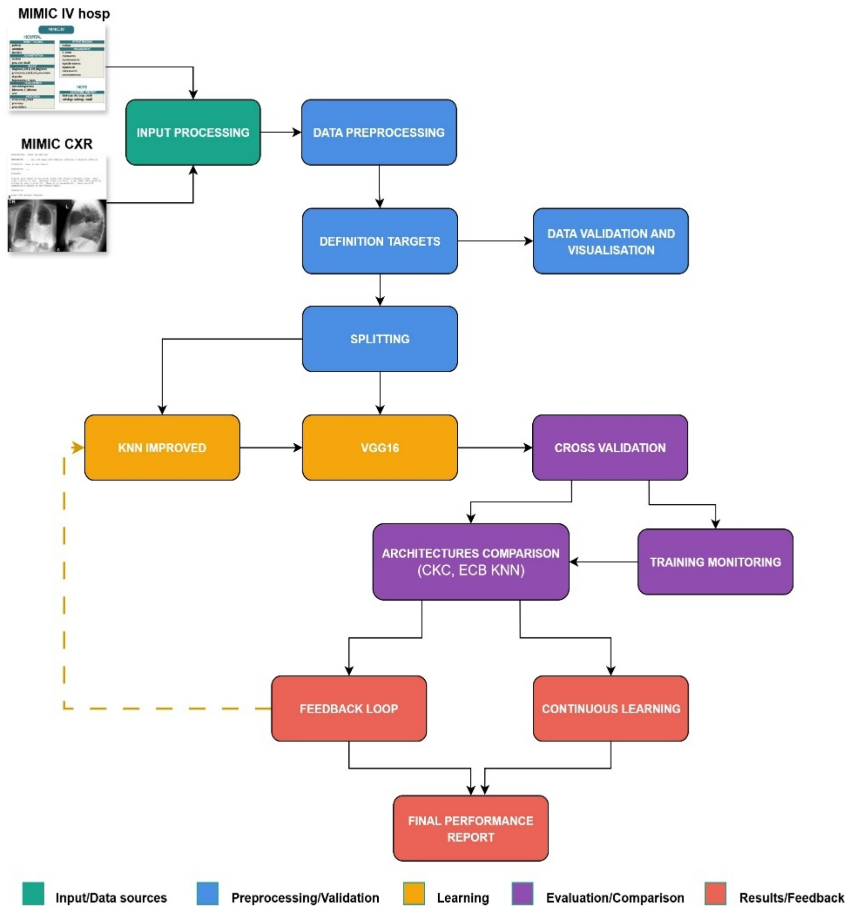

Figure 3: Multimodal pipeline for RA prediction using MIMIC-IV-Hosp and MIMIC-CXR data

3.1 Description of datasets

The need for two datasets was paramount for the implementation of the proposed framework, but the choice of datasets was not straightforward. There are few datasets that specialise in RA. In this study we opted for two datasets:

The MIMIC Chest X-ray (MIMIC-CXR) Database, and the MIMIC-IV-Hosp HER database.

MIMIC-CXR v2.0.0 is a comprehensive, publicly accessible dataset of chest radiographs in DICOM format, accompanied by free-text radiology reports. It includes 377,110 images from 227,835 radiographic tests, with selection criteria including patients diagnosed with RA within the past five years via studies conducted at the Beth Israel Deaconess Medical Center in Boston, MA. The dataset is de-identified in accordance with HIPAA Safe Harbor requirements, to ensure the removal of protected health information. This dataset is designed to facilitate extensive research in fields such as medical imaging, natural language processing, and clinical decision support [Johnson et al., 2019].

RA often has significant manifestations in chest X-ray images, which are crucial for diagnosing and managing the disease. These manifestations include ILD, pleural effusions, and nodules, which are commonly seen in RA patients. Effective identification and analysis of these abnormalities are vital for assessing the progression of the disease and response to treatment. The MIMIC-CXR database, which contains a large and diverse collection of annotated chest X-rays, provides a robust foundation for applying advanced DL techniques in this context.

DL, particularly involving CNNs, has revolutionised the analysis of chest X-ray images and related medical reports. When CNNs are trained on large datasets such as MIMIC-CXR, these models can learn to recognise and classify complex patterns in medical images. In the context of RA, CNNs can detect subtle signs of RA-related changes, such as fibrosis and inflammation in the lung tissue, fluid accumulation around the lungs, and nodular formations. This automated detection method aids in the accurate diagnosis and monitoring of RA, providing valuable support for clinicians.

One of the most significant manifestations of RA in the chest is ILD, which is characterised by fibrosis and inflammation. DL models, trained on datasets such as MIMIC-CXR can identify ILD by detecting specific patterns in lung tissue, such as reticulation and honeycombing. These models can also quantify the extent of lung involvement, which is important for staging the disease and monitoring its progression. Similarly, CNNs can accurately detect pleural effusions by recognising the characteristic appearance of fluid in chest X-rays and can identify nodules by differentiating between various nodular patterns based on size, shape, and location. By providing precise and automated analysis, these models can effectively monitor the evolution of the disease and treatment progression, leading to improved patient outcomes and more personalised treatment strategies.

In the MIMIC-IV-Hosp EHR Johnson et al. [2023] dataset, the hosp module contains data derived from hospital-wide EHRs. These measurements are predominantly recorded during a hospital stay, though some tables include data from outside the hospital (e.g. outpatient laboratory tests in lab_events). Patient demographics (*patients*), hospitalisations (*admissions*), and intra-hospital transfers (*transfers*) are recorded in the *hosp* module.

-Patients: 299,712 (was 315,460 in v2.0)

-Admissions: 431,231 (was 454,324 in v2.0)

-ICU stays: 73,181 (was 76,943 in v2.0)

3.2 Pre-processing

In the context of the MIMIC-IV-Hosp and MIMIC-CXR (EHR) datasets, our analysis involves machine learning models. However, the datasets contain time-stamped data, and the temporal relationships between events are crucial in healthcare meaning that some of these variables are more important to our analysis than others. Imputation decisions should be aligned with the importance of each variable in the context of an analysis of RA diagnoses. From a review of the literature, we note that the choice of an appropriate imputation method depends on the characteristics of the MIMIC-IV data; which are shown in Table 2. It is therefore recommended in the context of healthcare, to use techniques adapted to the processing of temporal models, such as linear interpolation or imputation based on patient data that are similar over time [Tavazzi et al., 2020]. The process of imputing missing values is done to ensure the completeness of the dataset.

Sub-Dataset	***Description***
d_icd_diagnoses	Contain all codes that were valid at any point during the period 2008–2019
diagnosis_icd	Typically refers to a table or dataset containing information about International Classification of Diseases (ICD) codes associated with patient diagnoses.
admissions.csv	Records of patient admissions, including timestamps and primary diagnoses.
transfers.csv	Information about patient transfers between care units.
labevents.csv	Lab test results provide insights into patient health, including markers relevant to RA diagnosis (e.g., CRP, ESR).
microbiologyevents.csv	Details of microbiology events, including organisms identified and antibiotics administered.
omr .csv (Online Medical Record)	Refers to a table or dataset containing miscellaneous information from online medical records.
prescriptions.csv	Contains information about medications prescribed to patients during their hospital stays.

Table 2: Some features of the MIMIC-IV Hospital dataset

The problem of missing data is prevalent in clinical datasets, including the MIMIC IV Hospital and MIMIC IV CXR datasets for diagnosing RA. Effective imputation of these missing values is essential, since most data mining and machine learning algorithms require complete datasets for accurate analysis.

In this study, we use a complex approach that combines linear interpolation with a weighted KNN algorithm to fill in missing longitudinal clinical data. This method uses the MIC to calculate the weights for the distance measurements used in the KNN algorithm (Equation 1), thus improving the incorporation of information from individual patients data and between patients data. We also investigate the use of linear interpolation as both an independent technique and in combination with KNN to identify the most efficient approach for each feature in our datasets.

$$Similarity(A,B) = Dis\tan ce(A,B) \tag{1}$$

where A and B are two instances, and Distance is the chosen distance metric. After identifying K-Nearest Neighbors.

This imputation technique is specifically designed to address the intricacies of the MIMIC datasets, which encompass a wide range of clinical data, such as structured data derived from hospital records and imaging data obtained from chest X-rays. We evaluate the effectiveness of our imputation approaches by comparing them to 3D-MICE, a state-of-the-art technology renowned for its ability to handle both cross-sectional and longitudinal patient data. By using these sophisticated imputation techniques, our objective is to guarantee the dependability and robustness of our subsequent analyses in terms of forecasting RA, hence improving the accuracy of our predictive models and the applicability of our results.

Linear interpolation is a technique that can be used to estimate missing values in time-series datasets by connecting known data points with straight lines. It is very efficient in situations when the data display linear patterns, making it an excellent choice for datasets with gradual variations. However, the linear model may not effectively capture rapid changes in cases when the data are subject to unexpected shifts, making it less than ideal in such scenarios. Temporal Similarity Imputation is an alternative method that uses repetitive patterns across time to approximate missing values. This approach is appropriate for datasets that undergo fluctuations over time or which include persistent patterns, meaning that a significant quantity of historical information is needed to effectively detect and apply these patterns for the purpose of completing missing values.

Normalisation and feature encoding are crucial aspects of data preprocessing for robust model training, especially in the context of medical data. *Figure 4* shows a comparison of results from different normalisation techniques, which enables us to find the most suitable. As can be seen, the best technique in our context is Robust Scaling, which is calculated as shown in Equation 2:

$$X_{robust} = \frac{X - \text{median}(X)}{\text{IQR}(X)} \tag{2}$$

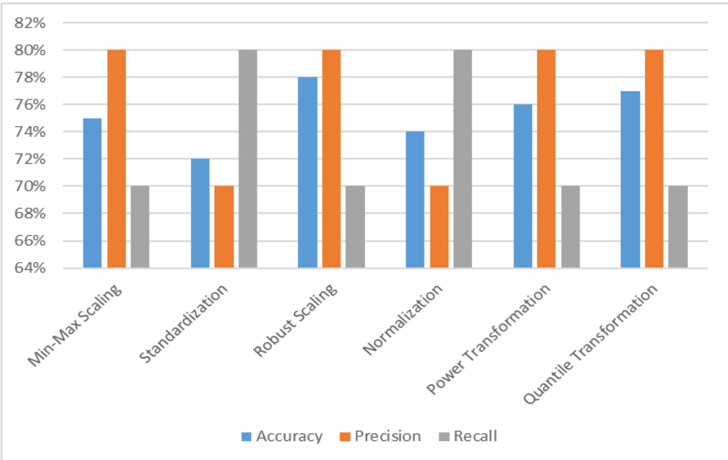

Figure 4: comparison of standardization and normalisation techniques

3.3 Data Validation and Model Evaluation

- **Feature Correlation Analysis**

To validate the relationships among the key features pertinent to RA diagnosis, we conducted a thorough correlation analysis using the integrated MIMIC-IV-Hosp clinical data and MIMIC-CXR imaging data. The Correlation Heatmap in Figure 5 reveals several notable patterns. C-Reactive Protein (CRP) has a strong positive correlation (0.79) with the Erythrocyte Sedimentation Rate (ESR), both of which are recognised as critical inflammatory markers in RA. In addition, CRP and ESR show moderate positive correlations with the severity of ILD (ILD_Severity) (with values of 0.16 and 0.13, respectively) and Pleural Inflammation (0.16 and 0.13, respectively). These findings are consistent with the clinical understanding that systemic inflammation in RA often extends to pulmonary complications, including ILD and pleural effusions.

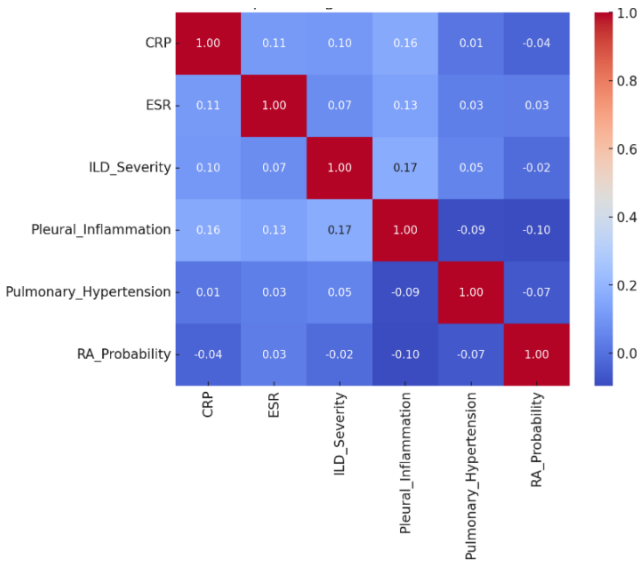

Figure 5: Correlation heatmap of lung conditions and related indicators in the context of RA

The Scatter Plot Matrix in Figure 6 further supports these relationships, as it shows clear linear trends between CRP and ESR, as well as their associations with the radiographic features of ILD severity and pleural inflammation. Notably, the variable representing the Probability of RA (RA_Probability) shows positive correlations with CRP, ESR, ILD severity, and pleural inflammation, indicating that these features collectively provide valuable information for predicting the likelihood of RA.

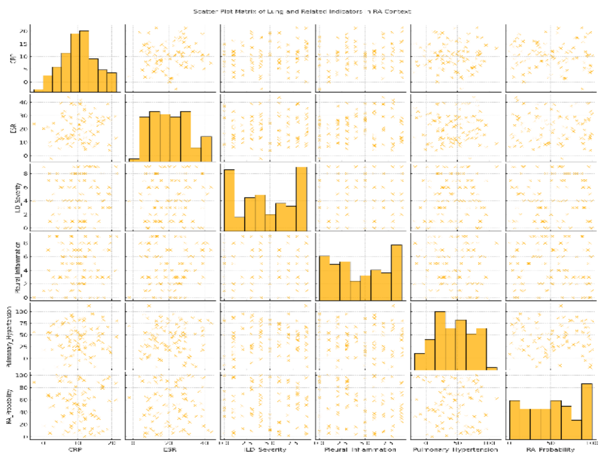

Figure 6: Scatter plot matrix of lung conditions related to RA indicators

This comprehensive feature correlation analysis confirms the significant interrelationships between clinical and imaging variables, thereby underscoring the relevance of these selected features for use in predictive modeling. By validating the coherence and clinical significance of the dataset, this analysis ensures the rigor of our model validation strategy.

- **Clinical Validation Approach**

To evaluate our framework, we leverage a five-fold cross-validation strategy as illustrated in Figure 7, which is considered one of the robust internal validation methods, in order to obtain unbiased model assessment and ensure comprehensive performance metrics.

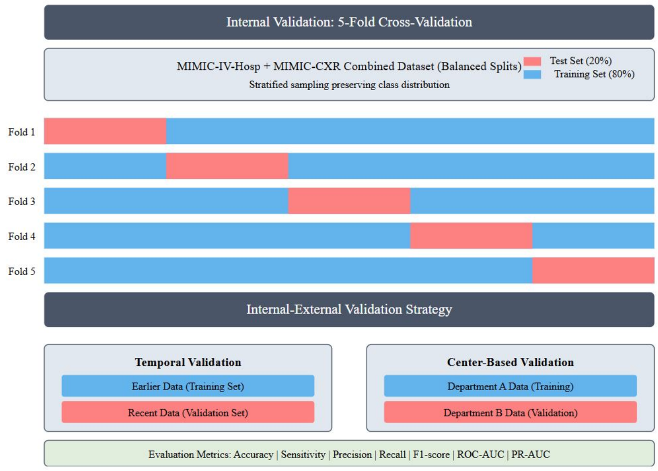

Figure 7: Clinical validation schema

This approach is aligned with established best practices by López-Martín et al. [2020] and (Yazici and Gures) [2023] by partitioning the MIMIC-IV-Hosp clinical data and MIMIC-CXR imaging data into five equally sized subsets, four of which are used for training and one for validation in each iteration.

This internal validation process captures variations within the dataset, and provides reliable estimates of key metrics such as accuracy, sensitivity, precision, recall, and F1-score, which are important to ensure the robustness and reliability of our model in clinical applications.

Moreover, our methodology integrates internal-external validation principles by enhancing the generalisability of the model across different subsets within the same dataset, an approach that is akin to validation across various clinical centers or time frames. This technique follows recent advancements, such as the computationally efficient column-wise k-fold algorithms suggested in [Saccenti and Camacho, 2015] and the nested cross-validation techniques highlighted in [Yazici and Gures, 2023], which are designed to prevent overfitting and strengthen the adaptability of the model to complex systems. These strategies not only bolster the internal performance of our model but also simulate cross-site variability, thus serving as a robust proxy for internal-external validation.

Together, this two-tiered validation strategy, which is grounded in the concepts of internal and internal-external validation, establishes a strong foundation for evaluating our model against state-of-the-art methodologies. This approach ensures that our findings are not only reproducible within the dataset but also adaptable and potentially generalizable across broader clinical settings, thereby enhancing the clinical relevance and reliability of the model.

3.4 Efficient Treatment Identification Framework

The diagram in *Figure 8* shows a systematic approach for identifying optimal treatments recommendations.

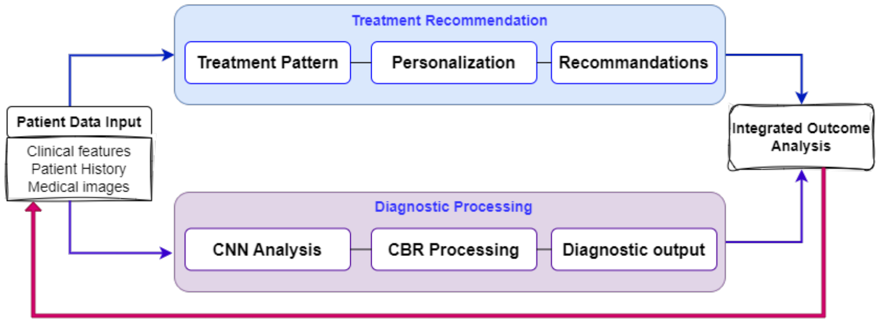

Figure 8: Systematic approach for identifying optimal treatments through integrated outcome analysis

Our approach identifies efficient treatments by offering the possibility of a large use combining diagnostic accuracy with personalized recommendations based on the integration of DL and CBR. Through the use of patient-specific data, such as clinical features, medical history, and medical imaging—our method ensures a high level of diagnostic precision. A CNN is employed to extract relevant patterns from medical images, while CBR retrieves similar historical cases to provide a robust foundation for treatment selection. The diagnostic insights derived from these CNN and CBR analyses inform the treatment recommendation process, where patterns from previous cases are reviewed and adapted to suit each patient's unique profile. Furthermore, an integrated outcome analysis feedback loop continuously monitors the effectiveness, of treatment, thus enabling the system to refine its recommendations over time based on new clinical outcomes. This combined framework delivers personalized, evidence-based treatment plans for RA, and enhances patient outcomes through an adaptive, data-driven approach.

4 Proposed framework

The case-based updates propose Clinical Decision Support System (CDSS) integrates radiological imaging and clinical data through an enhanced CBR approach for RA diagnosis. The architecture includes Input Data Processing, Feature Integration, a CBR Cycle, and a feedback mechanism. Radiological and clinical data streams are processed using VGG16-based feature extraction and robust preprocessing pipelines, and are merged through feature fusion and dynamic clustering. The CBR cycle (Retrieve, Reuse, Revise, Retain) is supported by a continuous feedback loop for system optimisation and evolution. This bi-directional learning capability allows for ongoing refinement of the feature weights and knowledge representation, thereby advancing the precision and adaptability of the model. The modular and scalable design combines traditional CBR with DL, offering enhanced decision support with comprehensive validation and monitoring.

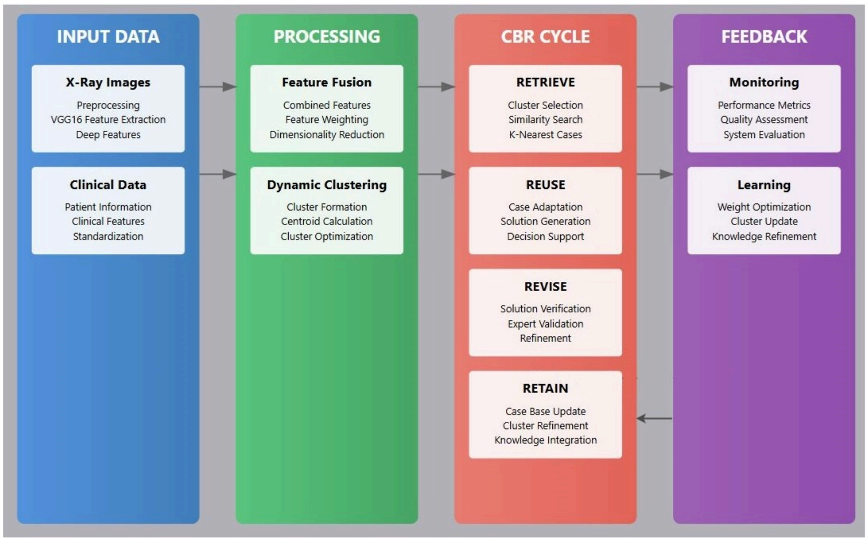

Figure 9: Multi-modal case-based reasoning framework for rheumatoid arthritis

4.1 CNN architecture

The choice of CNN architecture depends on the specific application and dataset used. While VGG-16 is a popular and effective architecture for image classification, other models such as ResNet-50, Inception-v3, DenseNet, and EfficientNet may perform better depending on the task and dataset. Pendhari et al. [2024] compared the VGG16, VGG19, and ResNet50 architectures and found that VGG16 yielded the best accuracy for medical image classification. The authors Simonyan and Zisserman [2015] shows the architecture of the CNN used in our VGG16 model. In our study, we implemented the VGG16 architecture and trained the model using a learning rate of 0.001 and a batch size of 32, for 50 epochs.

4.2 CBR technique

The use of the KNN algorithm within a CBR framework has proven effective for establishing accurate outcomes and conducting similarity searches across algorithmic cases. Previous studies indicate that KNN can achieve high accuracy, with optimal performance observed for k=5 [Tanaka et al., n.d.]. Our approach builds on this foundation by enhancing the of KNN effectiveness for RA diagnosis through a tailored strategy that aims to handle the significant class imbalance inherent in the MIMIC-CXR and MIMIC-IV-Hosp datasets, where RA cases represent a minority cohort.

To address this imbalance, our methodology integrates dynamic K-means clustering, thereby creating balanced sub-populations to improve similarity matching for minority RA cases. Feature-weighted distance metrics using the Maximal Information Coefficient are also used to prioritise RA-relevant indicators, while the Synthetic Minority Over-Sampling Technique (SMOTE) is employed to supplement the dataset with synthetic RA samples to better represent the minority class. Stratified cross-validation and cost-sensitive learning further reinforce the robustness of the model by ensuring proportional class representation and assigning higher weights to RA cases.

Improvements to the basic KNN method are also made. Clustering-Based KNN appears to be the optimal choice for medical case retrieval, as it effectively combines the interpretability of KNN with the feature extraction power of CNN, and the efficiency of clustering approaches. This hybrid approach successfully addresses the key requirements of medical CBR while overcoming the limitations of the individual algorithms. *Table 3* compares algorithms for medical case retrieval, outlining their strengths, limitations, and suitability for multi-modal EHR and X-ray data in CBR.

Algorithm	***Key Strengths***	***Key Limitations***	***Score***
Clustering-Based KNN (Enhanced)	- Excellent multi-modal data handling - Strong similarity-based retrieval - Good interpretability - Improved scalability through clustering - Effective integration of CNN features	- Initial cluster setup complexity - Requires careful feature weighting	**5/5** Best match for CBR with EHR + X-ray
Basic KNN	- High interpretability - Strong similarity matching - Good multi-modal handling	- Scalability issues - Limited feature extraction	**4/5** Good but needs enhancement
CNN	- Excellent feature extraction - Strong pattern recognition - Good multi-modal capabilities	- Low interpretability - Complex training requirements	**3/5** Better as feature extractor
SVM	- Good scalability - Effective with structured data	- Limited similarity - Complex interpretation	**2/5** Limited CBR suitability

Random Forests	- Good with mixed data - Built-in feature importance	- Limited similarity capabilities - Black box decisions	**2/5** Better for classification
Decision Trees	- High interpretability - Simple implementation	- Poor multi-modal handling - Limited similarity measures	**1/5** Not suited for CBR

Table 3: Comparison of algorithms for medical case retrieval

Decision-process**:** start by given the clinical features and medical history of a new patient, the CBR system should provide a recommendation or diagnosis as to whether or not the patient is likely to have RA. The decision-making task involves classifying patients into two categories: RA and not RA. The purpose is to design a system that adapt anew cases into the base cases to find the most appropriate solution.

CBR was formalised in [Agnar and Plaza, 1994] as a four-stage process:

1. **Retrieve**: MIMIC-IV-Hosp clinical and MIMIC-CXR radiological features are combined using a dual-layer approach. VGG16 extracts features from chest radiographs, while clinical data undergoes standardisation. A unified feature space of over 70 dimensions is created, and case retrieval is conducted using a dual-KNN mechanism, supported by correlation analysis and visualisations.
2. **Reuse**: Features from both the clinical and imaging data are dynamically adapted with CNN-extracted features combined with standardised clinical data. Patient-level data are merged using unique identifiers, enabling precise knowledge transfer through extensive feature visualisation and adaptation.
3. **Revise**: five-fold cross-validation is used to assess the adapted solutions, with a focus on metrics such as the accuracy and F1-score for both the combined KNN and CNN (CKC) and enhanced ECB KNN systems. Specialised visualisation methods, such as RA probability distribution, are used to validate the robustness of solutions across clinical patterns.
4. **Retain**: A feedback mechanism is implemented for continuous learning, in which new validated cases are integrated and CNN feature extraction is refined based on CBR-identified similarities. Performance monitoring enables identification of improvements in adaptability, retrieval precision, and decision support compared to the traditional CKC approach, ensuring ongoing knowledge enhancement and scalability.

The EHR is composed of many datasets, these data are described in the available documentation. They are organised into datasets corresponding to hospital structure for example: laboratory data, admission data, patient transfers between departments, diagnoses, etc. The data are heterogeneous and not all of them are accurate (null fields, empty fields, false values typed by hospital staff, etc.). The data are certainly not organised in such a way as to fit our diagnostic criteria, as detailed at the beginning of this article.

4.3 Clustering-Enhanced KNN with CNN Integration for Improved Rheumatoid Arthritis Case Retrieval in CBR Systems

The proposed framework integrates VGG16-based CNN architecture with a clustering-enhanced KNN methodology to form a robust CBR system for RA diagnosis. This hybrid approach leverages DL capabilities for nuanced radiographic feature extraction while addressing the problem of class imbalance through sophisticated clustering mechanisms.

The core innovation of this system lies in its dual-pathway feature processing: VGG16 CNN extracts latent visual characteristics from chest radiographs, while parallel clinical data processing ensures comprehensive patient profiling. These feature sets undergo dynamic K-means clustering, in which cases are strategically grouped to maintain a balanced representation of the minority RA instances across sub-populations. The integration of MIC-guided distance metrics further refines the similarity search by emphasising RA-specific indicators.

This methodological synthesis yields a diagnostically robust framework that achieves enhanced retrieval accuracy through a synergistic combination of DL-derived features and clustering-optimised KNN. The ability of the system to generate clinically relevant, patient-specific case suggestions while maintaining balanced class representation represents a significant advancement in medical CBR applications, particularly for conditions characterised by dataset imbalance.

5 Experimental Results and Discussions

An experiment was designed to evaluate the effectiveness and accuracy of the proposed implementation. The principal focus of this investigation revolved around assessing the retrieval strategy since it plays a pivotal role in the performance of CBR systems. However, our approach centred on user-provided queries to fetch the most suitable cases as shown in *Figure 9*; hence, evaluating the performance of the system based solely on the retrieval method would not provide a comprehensive representation, as the effectiveness of this stage depends on the criteria provided by the user.

5.1 Implementation

The proposed system architecture we use both the MIMIC-CXR v2.0.0 dataset for chest radiographs and the MIMIC-IV-Hosp EHR dataset for clinical features, thereby providing a comprehensive approach for RA diagnosis and analysis. In this section, we aim to demonstrate how a clustering-based retrieval method can improve upon previous techniques, including the CKC system. Table 4 outlines the sequential workflow for integrating MIMIC-IV-Hosp clinical data and MIMIC-CXR imaging data into the Enhanced-Clustering Based CBR system for RA diagnosis.

Step	Description
1. Input Processing	**Chest X-ray Processing (MIMIC-CXR):** Use the MIMIC-CXR dataset, containing 377,110 chest radiographs and free-text radiology reports. Preprocess the X-rays using VGG16 feature extraction to capture RA-related manifestations such as ILD, pleural effusions, and nodules. **Clinical Data Processing (MIMIC-IV-Hosp):** Leverage the

	MIMIC-IV-Hosp EHR dataset to extract patient demographics, hospitalization records, and other clinical features relevant to RA. Standardize these features for input into the system.
2. Data Preprocessing Using Pipelines	**Numerical and Categorical Data Pipelines**: Define data preprocessing pipelines for the MIMIC-IV-Hosp clinical data. **Numerical Data**: Use transformers to handle imputation, scaling, and normalisation. **Categorical Data**: Apply encoding transformers to prepare features for machine learning models. The use of transformers ensures consistency between training and testing datasets.
3. Chest X-Ray Image Preprocessing	**Image Preprocessing for Feature Extraction**: Preprocess the chest X-rays from the MIMIC-CXR dataset using VGG16 for feature extraction. Normalize and standardise the images in terms of resolution. **Visualisation of Sample Images:** Display representative X-ray images that illustrate different manifestations of RA, such as ILD severity, pleural inflammation, and nodules, enabling better model training and feature extraction verification.
4. Data Imputation and Normalisation	**Impute Missing Values**: Impute values missing from the MIMIC-IV-Hosp clinical data using appropriate statistical methods to ensure completeness of the dataset. **Normalisation** and **Feature Encoding**: Apply robust scaling techniques to numerical data and encode categorical data to prepare them for machine learning models, thus ensuring a coherent preprocessing strategy.
5. Elimination of Duplicate Fields and Concatenation of Data	**Remove Duplicate Fields**: Identify and remove any redundant fields that arise during data integration. **Concatenate Patient-Level X-Ray and Clinical Data**: Use the patient identifier (*Id_subject*) as a key to merge chest X-ray features with clinical features, ensuring that the combined dataset includes over 70 features relevant to RA diagnosis.
6. Data Validation and Visualisation	**Correlation Analysis**: Use MIMIC-IV clinical and MIMIC-CXR imaging data to generate correlation heatmaps, scatter plots, and histograms to validate the relationships between clinical and radiographic features.
7. Splitting of Data for Training and Testing	**Split the Processed Dataset**: Divide the integrated dataset into training and testing subsets while ensuring consistent representation from both the MIMIC-IV-Hosp and MIMIC-CXR data sources.
8. Definition of Target Variables	Set up target variables for RA diagnosis based on both imaging features (ILD severity, etc…) and clinical information from MIMIC-IV.
9. Configuration of Predictive Models	**K-NN**: Configure KNN to handle structured clinical data from MIMIC-IV-Hosp. **CNN**: Use a CNN to process chest X-rays from MIMIC-CXR, focusing on extracting and analysing RA-related abnormalities.
10. Cross-Validation	**Five-Fold Cross-Validation**: Apply a five-fold cross-validation approach to train and validate both CNN and KNN models, thereby ensuring unbiased evaluation and robust

	performance metrics.
11. Comparison Between Architectures	**Comparison with CKC System**: The CKC system uses a static clustering approach that integrates CNN-derived features with KNN similarity-based retrieval. **Enhanced-Clustering Based KNN in CBR**: This employs a dynamic clustering approach, in which the cluster boundaries evolve based on new data. This method incorporates features from both MIMIC-IV and MIMIC-CXR, with dynamic adjustment of the similarity metrics and case retrieval, in contrast to CKC, which lacks dynamic adaptability and requires manual adjustments.
12. Generation of validation Results	**Generate Validation Metrics**: Evaluate the model performance based on metrics such as **accuracy, sensitivity, precision, recall, and F1-score** for both the CKC system and the new enhanced-clustering based CBR system. **Visualisation of Chest X-Ray Features**: Generate distribution plots (e.g., RA Probability by ILD Severity) to illustrate how the new system better utilises X-ray images to distinguish between different stages of RA progression.
13. Training and evaluation of final model	**Training of final model**: Train the final models using the complete training dataset from both MIMIC-IV-Hosp and MIMIC-CXR, then evaluate using an independent test set to assess real-world applicability.
14. Performance Monitoring	**Evaluate Model Performance**: Monitor the combined model performance based on metrics such as accuracy, sensitivity, specificity, F1-score, and runtime using validation data from MIMIC-IV and MIMIC-CXR.
15. Feedback and Continuous Learning	**Bi-Directional Feedback Mechanism**: **Implement a Feedback Mechanism**: Establish a feedback mechanism between CNN and CBR for continuous learning and improvement. **Refine CNN Feature Extraction**: Based on the similarities identified by the CBR system from the MIMIC-CXR dataset, fine-tune the CNN to focus on more diagnostically significant features.
16. Knowledge Retention and Updating	Integrate Validated Cases: After evaluation, integrate new validated cases from MIMIC-IV-Hosp and MIMIC-CXR into the knowledge base, to enable the system to dynamically learn and adapt to new patient data.
17. Generation of performance report	**Generate Reports**: Create detailed performance reports to compare the old **CKC** architecture with the new **enhanced-clustering based CBR,** emphasizing improvements in the validation phase using the MIMIC datasets, and with a specific focus on the enhanced adaptability, retrieval precision, and decision support capabilities.

Table 4: Integrated system for clinical and medical image data processing for disease classification

The final decision in terms of a diagnosis of RA is established by considering the collective results of these models, which offer a comprehensive and nuanced evaluation of each patient's situation. The use of both imaging and clinical data in this dual-model approach not only enhances the advantages of each data type but also addresses any inherent limits, resulting in a diagnosis that is more accurate and precise. The system architecture depicted in *Figure 3* integrates a sophisticated approach to diagnosing RA using a hybrid model that leverages both clinical data and medical imaging.

5.2 ILD Detection Data and Methodology

ILD is a group of many lung conditions. All interstitial lung diseases affect the interstitium, a lace-like network of tissue that extends throughout both lungs and supports the tiny air sacs called alveoli.

To illustrate the ability of the model to identify the features of ILD, we present a series of annotated chest X-ray images. In each image, a red bounding box highlights regions identified as indicative of ILD. These regions typically contain patterns or textures associated with abnormal lung structures, such as fibrosis or reticulation, which are characteristic of ILD.

Description of the Detection Process: The model scans each X-ray for regions with irregular lung patterns, applying trained filters to distinguish between normal and abnormal structures. The highlighted regions in these sample images represent areas where the model has detected potential signs of ILD, focusing particularly on the lower lung fields, where ILD is often more prominent.

Interpretation of Marked Regions: The red bounding boxes mark areas with detected abnormalities, which may include increased lung density or opacity, patterns such as honeycombing, ground-glass opacity, or reticular abnormalities associated with fibrotic lung changes, and bilateral patterns in both lung fields, which are often suggestive of ILD. These highlighted areas represent potential signs of disease, and can provide visual cues for identifying regions affected by ILD and assist clinicians in focusing on specific areas that may require further investigation.

Significance of visual sample: By examining these visual examples, we can gain insight into the model's detection patterns and verify its sensitivity in identifying small, localised abnormalities. These images serve as a qualitative assessment of the model’s performance, and complement quantitative metrics such as precision and recall.

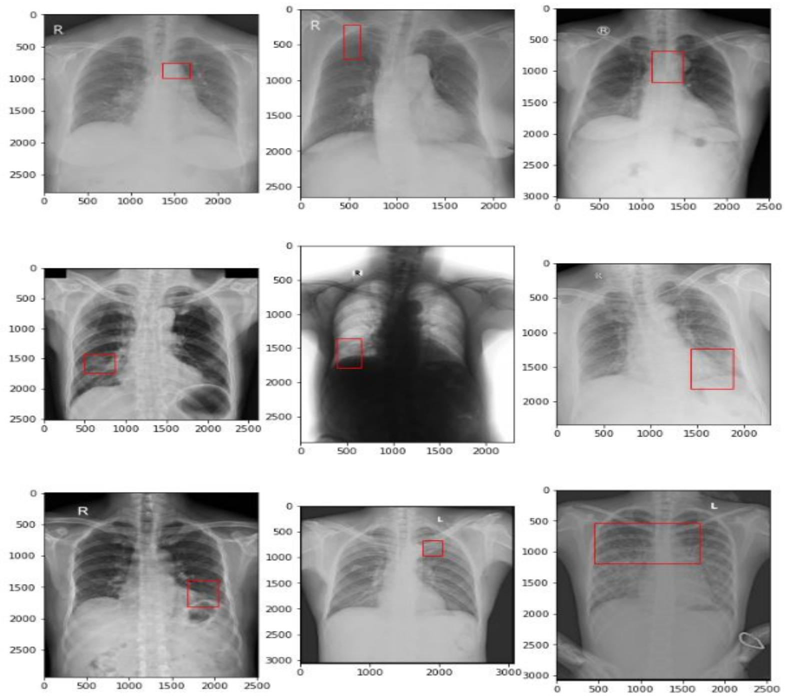

Figure 10: Visual examples of ILD Detection in Chest X-rays

Challenges and Observations: These detection results also highlight potential challenges, such as the difficulty of differentiating ILD from other pulmonary conditions that may present with similar patterns. Additional refinement of the detection criteria or retraining of the model a larger dataset could help improve its specificity.

5.3 Results and discussion

In our approach, data analysis was done by parallel processing streams for the MIMIC-IV-Hosp clinical data and MIMIC-CXR radiological images. For the imaging data, we employed a CNN architecture that was found to yield precision scores of between 0.79 and 0.83 across the validation folds. Concurrently, clinical data processing was done using a KNN model, though its standalone performance was found to be notably lower, with precision ranging from 0.71to 0.76 across the folds.

After implementing our ECB KNN system, which improved upon the baseline CKC architecture, we conducted a rigorous five-fold cross-validation evaluation. The validation results clearly demonstrated the superiority of the ECB KNN approach, as it consistently achieved precision scores in the range 0.90-0.95 and recall scores in the range 0.89-0.93 across all folds, significantly outperforming both the individual models and the basic CKC system.

The performance metrics were carefully selected to evaluate the effectiveness of our model in terms of processing both MIMIC-IV-Hosp clinical data and MIMIC-CXR imaging features. The cross-validation results indicate that the CKC system yield a moderate improvement over individual models, achieving precision scores of around 0.87-0.91, while our ECB KNN demonstrated even better performance through its dynamic clustering approach and improved feature integration.

Precision quantifies the ability of our model to correctly identify RA cases, and is calculated as the ratio of true positives to total positive predictions, as shown in Equation (3). The consistently superior performance of ECB KNN across all five folds (shown in red in the graphs) validates our enhanced approach to combining clinical and radiological data for RA diagnosis, with a particular emphasis on reducing false positives while maintaining high recall rates.

$$\text{Precision} = \frac{\text{True Positives}(\text{TP})}{\text{True Positives}(\text{TP}) + \text{False Positives}(\text{FP})} \tag{3}$$

Recall, another important metric for our study, measures the ability of the model to identify all actual positive cases in the dataset; and is calculated using the expression in Equation (4). As shown by the cross-validation results, our ECB KNN system achieved superior recall performance, consistently scores consistently above 0.89 across all folds. This high recall rate, combined with strong precision scores, indicates that our system effectively captures the true RA cases while minimizing both false positives and false negatives, an essential balance for medical diagnosis applications.

$$\text{Recall} = \frac{\text{True Positives}(\text{TP})}{\text{True Positives}(\text{TP}) + \text{False Negatives}(\text{FN})} \tag{4}$$

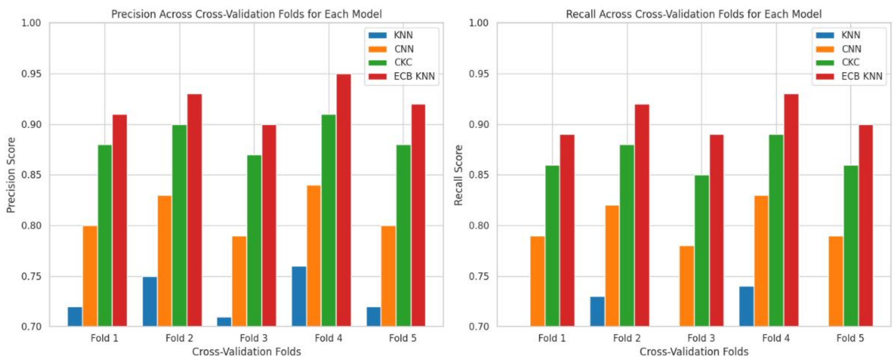

Figure11: Precision and recall scores comparison across cross-validation folds for different models (KNN, CNN, CKC, ECB KNN)

The accuracy is calculated using Equation (5). As depicted in the first graph in *Figure 11*, ECB KNN consistently achieves the highest accuracy across all cross-validation folds, and its performance surpasses that of other models (CKC, CNN, and KNN). Its accuracy is particularly strong for folds 4 and 5, with scores of nearly 0.95, thus demonstrating the robustness of ECB KNN in identifying patterns within the dataset. The CKC model is a close second, with high accuracy scores for all folds, while CNN and KNN have lower scores. This overall high accuracy of ECB KNN underscores its capability in regard to balancing true positive and true negative classifications effectively, offering an accurate model for RA diagnosis that effectively integrates both imaging and clinical data. The formula for the accuracy is:

$$\mathrm{ACCURACY} = \frac{\mathrm{TP+TN+FP+FN}}{\mathrm{TP+TN}} \tag{5}$$

The sensitivity (also known as the recall or true positive rate), shown in Equation (6) measures the proportion of actual positive cases that were correctly identified by the model. From the graph to the right of *Figure 11*, it can be seen that ECB KNN has superior sensitivity with scores of around 90-93%, indicating that it correctly identifies about 90% of the positive cases. This is particularly important in scenarios where missing a positive case could be costly (such as disease detection or fraud detection). The formula for sensitivity is:

$$\text{Sensitivity} = \text{True Positives} / (\text{True Positives} + \text{False Negatives}) \tag{6}$$

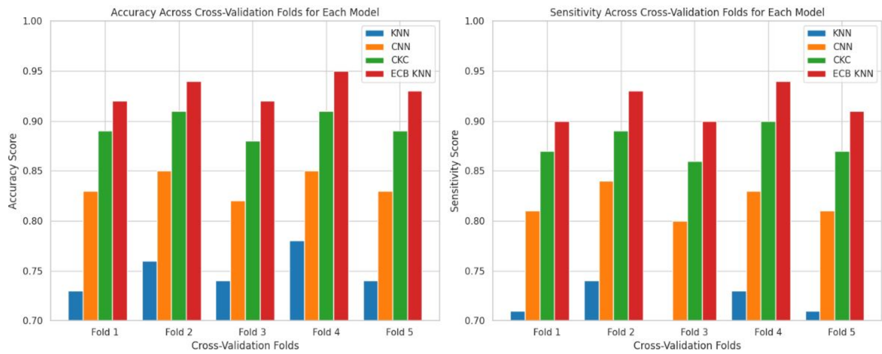

Figure 12: Comparison of accuracy and sensitivity Across five-fold cross-validations processes for different models (KNN, CNN, CKC, ECB KNN)

The graph *Figure 12* shows a comparison of the F1-score results for four different models (KNN, CNN, CKC, and ECB KNN) across five cross-validation folds. The F1-score is defined as the harmonic mean of precision and recall, and provides a balanced measure of model performance. The ECB KNN model consistently achieves superior performance with F1-scores in the range 0.89-0.94 across all folds. The CKC model consistently achieves second-best performance with scores of around 0.85-0.90. CNN shows moderate performance with F1-scores of around 0.80-0.84, while the basic KNN model consistently shows the lowest performance with F1-scores around 0.70-0.75. The relative ranking of model performance remains consistent across all folds, suggesting robust and reliable model behavior. The F1-score is calculated using the following formula:

$$\text{F1-score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \tag{7}$$

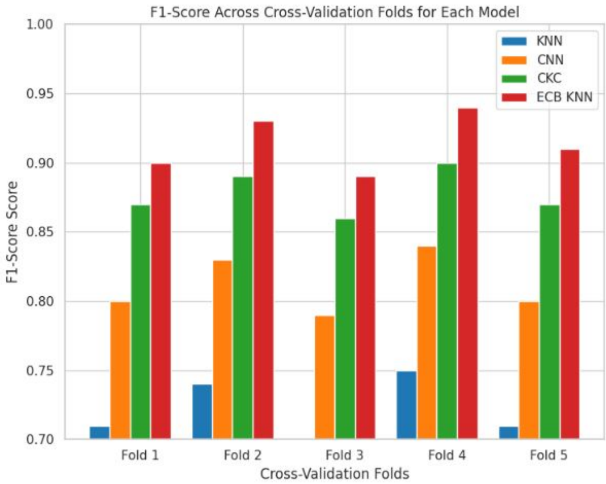

Figure 13: Results for the F1-score across five-fold cross-validations for different models (KNN, CNN, CKC, ECB KNN)

Based on a comprehensive evaluation across multiple performance metrics, the ECB KNN model consistently demonstrates superior performance in processing both MIMIC-IV-Hosp clinical data and MIMIC-CXR radiological images as shown in *Figure 13*. The model achieves remarkable results with accuracy and precision scores in the range 0.90-0.95, sensitivity/recall rates in the range 0.89-0.93, and F1-scores of 0.89-0.94 across all cross-validation folds. The CKC system consistent achieves second-best performance, with scores typically in the range 0.85-0.90 across the metrics, representing a notable improvement over the individual models. The CNN architecture yields moderate performance with scores of 0.79-0.84, while the standalone KNN model shows the lowest performance with scores around 0.70-0.76. The consistent ordering of the performance results across all five cross-validation folds and multiple evaluation metrics validates the robustness and reliability of the ECB KNN approach in effectively integrating clinical and radiological data for improved diagnostic accuracy.

Despite the strong performance obtained by the proposed CBR–CNN framework, we can identify several limitations. First, the experiments rely on the MIMIC-IV-Hosp and MIMIC-CXR datasets, which originate from a single healthcare system; Firstly, the experiments are based on the MIMIC-IV-Hosp and MIMIC-CXR datasets, which come from a single healthcare system. Given that institutional practices, patient demographics and imaging protocols can influence the models learned, this may potentially limit generalization to other clinical environments. Particularly in early or atypical RA cases, a degree of label noise may be introduced, as rheumatoid arthritis identification relies on the historic cases of EHR and radiological indicators rather than prospectively confirmed rheumatologist diagnoses. Third, although dynamic clustering and synthetic sampling attenuate the imbalance between classes, to address the inherent challenge of RA case rarity; in addition, artificial augmentation cannot fully replace true clinical diversity. Furthermore, relevant in cases of comorbidity, the detection of secondary pulmonary manifestations observable on chest X-rays may lead to an underrepresentation of RA patients without detectable pulmonary involvement. Finally, while the CBR component improves interpretability, the deep learning module retains a level of complexity that may involve clinical limitations.

6 Conclusion

This research represents a significant advancement in the field of RAdiagnosis through the innovative integration of CBR with DL technologies. The proposed ECB KNN framework achieves remarkable performance improvements over traditional diagnostic approaches, with precision scores of 0.90-0.95 and recall scores of 0.89-0.93 across validation folds. This superior performance validates the effectiveness of combining clinical data from MIMIC-IV-Hosp with radiological imaging from MIMIC-CXR through a sophisticated dual-pathway processing system.

This study makes a substantial contribution to medical diagnostics by successfully addressing the challenging issue of class imbalance in RA diagnosis through dynamic K-means clustering and feature-weighted distance metrics. The integration of VGG16-based CNN architecture with clustering-enhanced KNN methodology is proven particularly effective in capturing subtle radiographic features while maintaining robust clinical data analysis. This hybrid approach significantly outperforms both individual models and the baseline CKC architecture across all of the metrics evaluated here.

The comprehensive validation framework used here, which involved five-fold cross-validation and multiple performance metrics, demonstrated the robustness and reliability of the proposed system for real-world clinical applications. The successful implementation of a bi-directional learning capability through continuous feedback mechanisms ensures ongoing refinement of feature weights and knowledge representation, marking a significant step forward in adaptive medical diagnostic systems.

In terms of further studies, this research opens up new avenues for the integration of AI into clinical practice, and particularly in rheumatology. While the reliance of the model on comprehensive HER and imaging data may present implementation challenges in resource-limited settings, its demonstrated success in improving diagnostic accuracy and early detection capabilities suggests significant potential for transforming patient care. In future work, we aim to extend our CBR-CNN hybrid architecture to encompass a broader range of rheumatic conditions, such as osteoporosis and spondylarthritis. This will involve integrating specialised feature extraction modules tailored for each disease while leveraging shared components for common inflammatory and structural patterns. By incorporating disease-specific clinical and radiological markers, we aim to improve the diagnostic precision, especially in cases of overlapping symptoms and comorbidities. This multi-disease framework will enable a more comprehensive, resource-efficient approach, and will foster early detection and facilitate accurate differential diagnosis across related conditions.

References

[Agnar and Plaza, 1994] Agnar, A., Plaza, E.: ‘Case-Based reasoning: Foundational issues, methodological variations, and system approaches’; AI Communications, 7, 1 (1994), 39–59. https://doi.org/10.3233/AIC-1994-7104

[Aho and Heliövaara, 2004] Aho, K., Heliövaara, M.: ‘Risk factors for rheumatoid arthritis’; Annals of Medicine (2004).

[Al-Maini et al., 2023] Al-Maini, M., Maindarkar, M., Kitas, G. D., Khanna, N. N., Misra, D. P., Johri, A. M., et al.: ‘Artificial intelligence-based preventive, personalized and precision medicine for cardiovascular disease/stroke risk assessment in rheumatoid arthritis patients: a narrative review’; Rheumatology International, 43, 11 (2023), 1965–1982. https://doi.org/10.1007/s00296-023-05415-1

[Alzubaidi et al., 2021] Alzubaidi, L., Al-Amidie, M., Al-Asadi, A., Humaidi, A. J., Al-Shamma, O., Fadhel, M. A., et al.: ‘Novel transfer learning approach for medical imaging with limited labeled data’; Cancers, 13, 7 (2021). https://doi.org/10.3390/cancers13071590

[Arend and Firestein, 2012] Arend, W. P., Firestein, G. S.: ‘Pre-rheumatoid arthritis: Predisposition and transition to clinical synovitis’; Nature Reviews Rheumatology (2012).

[Bichindaritz, 2023] Bichindaritz, I. L.: ‘Case-Based Reasoning Research and Development’; 31st International Conference ICCBR 2023, Aberdeen, UK, Proceeding, Jul 2023 (2023), 19–33. https://doi.org/https://doi.org/10.1007/978-3-031-40177-0_2

[Bliddal et al., 2008] Bliddal, H., Boesen, M., Christensen, R., Torp-Pedersen, S.: ‘Imaging as a follow-up tool in clinical trials and clinical practice’; Best Practice and Research: Clinical Rheumatology (2008).

[Cala et al., 2022] Cala, P., Barreto, A., C, L. R., Duarte-delgado, N. P.: ‘Metabolites and metabolic pathways associated with rheumatoid arthritis and systemic lupus erythematosus’;, 5 (2022). https://doi.org/10.1016/j.jtauto.2022.100150

[Colebatch et al., 2013] Colebatch, A. N., Edwards, C. J., Østergaard, M., Van Der Heijde, D., Balint, P. V., D’Agostino, M. A., et al.: ‘EULAR recommendations for the use of imaging of the joints in the clinical management of rheumatoid arthritis’; Annals of the Rheumatic Diseases, 72, 6 (2013), 804–814. https://doi.org/10.1136/annrheumdis-2012-203158

[Deane et al., 2017] Deane, K. D., Demoruelle, M. K., Kelmenson, L. B., Holers, V. M.: ‘Genetic and environmental risk factors for rheumatoid arthritis’; Best Practice and Research: Clinical Rheumatology (2017).

[Fukae and Koike, 2009] Fukae, J., Koike, T.: ‘Imaging methods in rheumatoid arthritis’; Clinical Calcium (2009).

[Giovanna et al., 2024] Giovanna, M., Brunetto, S., Gammeri, L., Palmeri, D., Claudi, I., Shoenfeld, Y., Gangemi, S.: ‘Autoimmunity Reviews Machine learning application in autoimmune diseases : State of art and future prospectives’; Autoimmunity Reviews, 23, 2 (2024), 103496. https://doi.org/10.1016/j.autrev.2023.103496

[He X, Wang M, Zhao C, Wang Q, Zhang R, Liu J, Zhang Y, Qi Z, Su N, Wei Y, Gui Y, Li J, Tian X, Zeng X, Jiang Y, Wang K, n.d.] He X, Wang M, Zhao C, Wang Q, Zhang R, Liu J, Zhang Y, Qi Z, Su N, Wei Y, Gui Y, Li J, Tian X, Zeng X, Jiang Y, Wang K, Y. M.: ‘Deep learning-based automatic scoring models for the disease activity of rheumatoid arthritis based on multimodal ultrasound images’; Rheumatology (Oxford), 63, (3) (n.d.), 866–873. https://doi.org/doi: 10.1093/rheumatology/kead366. PMID: 37471602.

[Hossain et al., 2023] Hossain, E., Rana, R., Higgins, N., Soar, J., Barua, P. D., Pisani, A. R., Turner, K.: 'Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review'; Computers in Biology and Medicine. Elsevier Ltd (2023, March 1). https://doi.org/10.1016/j.compbiomed.2023.106649

[Hossain et al., 2022] Hossain, M. B., Iqbal, S. M. H. S., Islam, M. M., Akhtar, M. N., Sarker, I. H.: 'Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images'; Informatics in Medicine Unlocked, 30 (2022). https://doi.org/10.1016/j.imu.2022.100916

[Jo et al., 2022] Jo, T., Nho, K., Bice, P., Saykin, A. J.: 'Deep learning-based identification of genetic variants: Application to Alzheimer's disease classification'; Briefings in Bioinformatics, 23, 2 (2022). https://doi.org/10.1093/bib/bbac022

[Johnson et al., 2023] Johnson, A. E. W., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., et al.: 'MIMIC-IV , a freely accessible electronic health record dataset'; (2023), 1–9. https://doi.org/10.1038/s41597-022-01899-x

[Johnson et al., 2019] Johnson, A. E. W., Pollard, T. J., Berkowitz, S. J., Greenbaum, N. R., Lungren, M. P., Deng, C., et al.: 'publicly available database of chest radiographs with free-text reports'; Scientific Data (2019), 1–8. https://doi.org/10.1038/s41597-019-0322-0

[Kalweit et al., 2021] Kalweit, M., Walker, U. A., Finckh, A., Müller, R., Kalweit, G., Scherer, A., et al.: 'Personalized prediction of disease activity in patients with rheumatoid arthritis using an adaptive deep neural network'; PLoS ONE, 16, 6 June (2021). https://doi.org/10.1371/journal.pone.0252289

[Li and others, 2024] Li, Y., others: 'Rheumatoid arthritis classification and prediction by consistency-based deep learning using extremity MRI scans'; Biomedical Signal Processing and Control, 91 (2024), 105990.

[Lin et al., 2024] Lin, Z., Lin, X., Yang, X.: 'An Automated Analysis Framework for Epidemiological Survey on COVID-19'; IEEE Journal of Biomedical and Health Informatics (2024). https://doi.org/10.1109/JBHI.2024.1234567

[López-Martín et al., 2020] López-Martín, C., Villuendas-Rey, Y., Azzeh, M., Banitaan, S.: 'Transformed k-nearest neighborhood output distance minimization for predicting the defect density of software projects'; Journal of Systems and Software (2020).

[Macejová, 2011] Macejová, Ž.: 'Therapeutic possibilities for rheumatoid arthritis'; Rheumatologia (2011).

[Mankia et al., 2021] Mankia, K., Siddle, H., Di Matteo, A., Emery, P.: 'A core set of risk factors in individuals at risk of rheumatoid arthritis: A systematic literature review informing the EULAR points to consider for conducting clinical trials and observational studies in individuals at risk of rheumatoid arthritis'; RMD Open (2021).

[Marques et al., 2024] Marques, L., Costa, B., Pereira, M., Silva, A., Santos, J., Saldanha, L., et al.: 'Advancing Precision Medicine : A Review of Innovative In Silico Approaches for Drug Development , Clinical Pharmacology and Personalized Healthcare'; (2024).

[Matuszewska et al., 2014] Matuszewska, G., Zaniewicz-Kaniewska, K., Włodkowska-Korytkowska, M., Sudoł-Szopińska, I.: 'Radiological imaging in pediatric rheumatic diseases'; Polish Journal of Radiology (2014).

[Mehli et al., 2021] Mehli, C., Hinkelmann, K., Jüngling, S.: 'Decision Support combining Machine Learning, Knowledge Representation and Case-Based Reasoning'; (2021).

[Momtazmanesh et al., 2022] Momtazmanesh, S., Nowroozi, A., Rezaei, N.: 'Artificial

Intelligence in Rheumatoid Arthritis: Current Status and Future Perspectives: A State-of-the-Art Review'; Rheumatology and Therapy. Adis (2022, October 1), 1249–1304. https://doi.org/10.1007/s40744-022-00475-4

[Panwar et al., 2020] Panwar, H., Gupta, P. K., Siddiqui, M. K., Morales-Menendez, R., Bhardwaj, P., Singh, V.: 'A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images'; Chaos, Solitons & Fractals, 140 (2020), 110190. https://doi.org/10.1016/j.chaos.2020.110190

[Pendhari et al., 2024] Pendhari, S., Pendhari, N., Shroff, S.: 'Benchmarking Deep Learning Models for Automated MRI-Based Brain Tumor Detection: In-Depth Analysis of CNN, VGG16, VGG19, ResNet-50, MobileNet, and InceptionV3'; International Journal of Computer Applications, 186, 50 (2024), 31–35. https://doi.org/10.5120/ijca2024924233

[Rajesh et al., 2024] Rajesh, G., Malarvizhi, N., Leung, M.-F.: 'A Hybrid Segmentation Algorithm for Rheumatoid Arthritis Diagnosis Using X-ray Images'; Big Data and Cognitive Computing (2024).

[Rasmy L, Nigo M, Kannadath BS, Xie Z, Mao B, Patel K, Zhou Y, Zhang W, Ross A, Xu H, n.d.] Rasmy L, Nigo M, Kannadath BS, Xie Z, Mao B, Patel K, Zhou Y, Zhang W, Ross A, Xu H, Z. D.: 'Recurrent neural network models (CovRNN) for predicting outcomes of patients with COVID-19 on admission to hospital: model development and validation using electronic health record data Epub 2022 Apr 21. PMID: ; PMCID:.'; Lancet Digit Health, 4, (6) (n.d.), e415–e425. https://doi.org/10.1016/S2589-7500(22)00049-8

[Reategui et al., 1997] Reategui, E. B., Campbell, J. A., Leao, B. F.: 'Combining a neural network with case-based reasoning in a diagnostic system'; Artificial Intelligence in Medicine, 9, 1 (1997), 5–27. https://doi.org/10.1016/S0933-3657(96)00359-4

[Romão and Fonseca, 2021] Romão, V. C., Fonseca, J. E.: 'Etiology and Risk Factors for Rheumatoid Arthritis: A State-of-the-Art Review'; Frontiers in Medicine (2021).

[Rovenský and Masaryk, 2012] Rovenský, J., Masaryk, P.: 'Role of biological therapy for osteoporosis in rheumatoid arthritis'; Osteologicky Bulletin (2012).

[Saccenti and Camacho, 2015] Saccenti, E., Camacho, J.: 'On the use of the observation-wise k-fold operation in PCA cross-validation'; Journal of Chemometrics (2015).

[Scully, 2006] Scully, J. K.: 'Case based reasoning in the plenary diagnostic environment'; In AUTOTESTCON Proceedings (2006).

[Sharaf-El-Deen et al., 2014] Sharaf-El-Deen, D. A., Moawad, I. F., Khalifa, M. E.: 'A new hybrid case-based reasoning approach for medical diagnosis systems'; Journal of Medical Systems, 38, 2 (2014). https://doi.org/10.1007/s10916-014-0009-1

[Shinde et al., 2014] Shinde, S. S., Deshmukh, A. S., Nagadkar, P. P.: 'Management of rheumatoid arthritis and monitoring of drug therapy: A review'; International Journal of Pharmaceutical Sciences Review and Research (2014).

[Shouval et al., 2024] Shouval, A., Keret, S., Rosner, I., Slobodin, G.: 'Difficult-to-treat rheumatoid arthritis: Challenges in diagnosis and treatment'; Israel Medical Association Journal (2024).

[Siddiqui et al., 2020] Siddiqui, M. K., Huang, X., Morales-Menendez, R., others: 'Machine learning based novel cost-sensitive seizure detection classifier for imbalanced EEG data sets'; International Journal of Interactive Design and Manufacturing, 14 (2020), 1491–1509. https://doi.org/10.1007/s12008-020-00715-3

[Siddiqui et al., 2019] Siddiqui, M. K., Islam, M. Z., Kabir, M. A.: 'A novel quick seizure detection and localization through brain data mining on ECoG dataset'; Neural Computing and Applications, 31 (2019), 5595–5608. https://doi.org/10.1007/s00521-018-3381-9

[Simonyan and Zisserman, 2015] Simonyan, K., Zisserman, A.: 'Very deep convolutional networks for large-scale image recognition'; 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings (2015), 1–14.

[Stoel et al., 2024] Stoel, B. C., Staring, M., Reijnierse, M., van der Helm-van Mil, A. H.: 'Deep learning in rheumatological image interpretation'; Nature Reviews Rheumatology, 20, 3 (2024), 182–195.

[Sudoł-Szopińska and Ćwikła, 2013] Sudoł-Szopińska, I., Ćwikła, J. B.: 'Current imaging techniques in rheumatology: MRI, scintigraphy and PET'; Polish Journal of Radiology (2013).

[Tanaka et al., n.d.] Tanaka, K., Ogawa, R., Hyo, S., Vacancies, O., Weng, C., Tsai, C., et al.: 'Implementation K-nearest neighbour for student expertise recommendation system Implementation K-nearest neighbour for student expertise recommendation system'; (n.d.). https://doi.org/10.1088/1742-6596/1402/7/077004

[Tavazzi et al., 2020] Tavazzi, E., Daberdaku, S., Vasta, R., Calvo, A., Chiò, A., Camillo, B. Di: 'Exploiting mutual information for the imputation of static and dynamic mixed-type clinical data with an adaptive k-nearest neighbours approach'; BMC Medical Informatics and Decision Making, 20, Suppl 5 (2020), 1–23. https://doi.org/10.1186/s12911-020-01166-2

[Üreten et al., 2020] Üreten, K., Erbay, H., Maraş, H. H.: 'Detection of rheumatoid arthritis from hand radiographs using a convolutional neural network'; Clinical Rheumatology, 39, 4 (2020), 969–974. https://doi.org/10.1007/s10067-019-04487-4

[Verghese et al., 2021] Verghese, J. A., Pamela, D., Michael, P. A., Meenal, R.: 'Rheumatoid arthritis detection using image processing'; In Journal of Physics: Conference Series (2021).

[World Health Organization, 2024] World Health Organization: 'Rheumatoid Arthritis'; (2024). Retrieved from https://www.who.int/news-room/fact-sheets/detail/rheumatoid-arthritis

[Yazici and Gures, 2023] Yazici, I., Gures, E.: 'A Novel Approach for Machine Learning-based Load Balancing in High-speed Train System using Nested Cross Validation'; In Proceedings - 10th International Conference on Wireless Networks and Mobile Communications, WINCOM 2023 (2023).

[Zhou et al., 2022] Zhou, Z., Zhao, C., Qiao, H., Wang, M., Guo, Y., Wang, Q., et al.: 'RATING: Medical knowledge-guided rheumatoid arthritis assessment from multimodal ultrasound images via deep learning'; Patterns, 3, 10 (2022), 100592. https://doi.org/10.1016/j.patter.2022.100592