Liu *et al.*
Health Information Science and Systems (2023) 11:44
https://doi.org/10.1007/s13755-023-00246-7

Health Information Science and Systems

RESEARCH

Medimatrix: innovative pre-training of grayscale images for rheumatoid arthritis diagnosis revolutionises medical image classification

Linchen Liu[1], Yiyang Zhang[2] and Le Sun[2*]

Abstract

Efficient and accurate medical image classification (MIC) methods face two major challenges: (1) high similarity between images of different disease classes; and (2) generating large medical image datasets for training deep neural networks is challenging due to privacy restrictions and the need for expert ground truth annotations. In this paper, we introduce a novel deep learning method called pre-training grayscale images with supervised learning for MIC (MediMatrix). Instead of pre-training on color ImageNet, our approach uses MediMatrix on grayscale ImageNet. To improve the performance of the network, we introduce ShuffleAttention (SA), a self-attention mechanism. By combining SA with the multiple residual structure (ResSA block) and replacing short-cut connections with dense residual connections between corresponding layers (densepath), our network can dynamically adjust channel attention weights and receive image inputs of different sizes, resulting in improved feature representation and better discrimination of similarities between different categories. MediMatrix effectively classifies X-ray images of rheumatoid arthritis (RA), enabling efficient screening without the need for expert analysis or invasive testing. Through extensive experiments, we demonstrate the superiority of MediMatrix over state-of-the-art methods and that color is not critical for rich natural image classification. Our results highlight the potential of computer-aided diagnosis combined with MediMatrix as a valuable screening tool for early detection and intervention in RA.

Keywords: Medical image classification, Rheumatoid arthritis, Grayscale images, Pre-training, Computer-aided diagnosis, Deep learning

Introduction

Rheumatoid arthritis (RA) is an autoimmune disease that causes discomfort in small joints and has the potential to affect other limb joints due to immune system dysfunction [1]. However, traditional random forest classification methods (e.g., support vector machines [2] and random forest [3]) have limitations and limited recent improvements, being time-consuming and inconsistent across objects. Moreover, constructing large medical imaging datasets from scratch is challenging due to privacy constraints and the need for expert ground truth [4]. Tschandl et al. [5] presented a large-scale dataset for skin lesion categorization algorithms.

Imaging studies (e.g., X-rays) are essential in the diagnosis of RA and reveal important symptoms of RA (e.g., synovitis, joint space narrowing, joint effusion, and bone stiffness) [6]. Deep neural networks (DNNs), especially convolutional neural networks (CNNs), are widely used in medical image classification (MIC) and have shown impressive performance [7, 8]. Researchers have explored various diagnostic models based on CNNs to diagnose RA [9]. These models have shown promising results in detecting bone lesions and erosions and grading symptoms [10].

*Correspondence: lesun1@nuist.edu.cn
[2] Department of Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science and Technology, Nanjing 210044, China
Full list of author information is available at the end of the article

Two common training methods for CNNs are widely used: (1) training a model with randomly initialized weights; and (2) pre-training a model on a similar task and then fine-tuning it on the target task. However, existing deep learning-based diagnostic models often focus on detecting deep lesion features by deepening the network, which may result in the loss of original features [11].

In recent years, researchers have focused on studying automatic joint space quantification for RA diagnosis [2]. Yet, accurately monitoring early-stage joint damage and bone degradation in RA remains challenging due to limited accuracy in prior studies [6]. Our work addresses this by enhancing algorithm sensitivity and accuracy to the sub-pixel level. This advancement enables rheumatologists to precisely track RA evolution in its early stages annually.

In this paper, we introduce a novel network architecture called *pre-training of grayscale images with supervised learning for MIC (MediMatrix)* for accurate and efficient diagnosis of RA from X-ray images. Inspired by a variant model of U-Net [12], since medical images are grayscale, we transform the ImageNet dataset into grayscale images and pre-train them using MediMatrix, followed by fine-tuning with the obtained pre-training weights. The entire procedure is illustrated in Fig. 1. MediMatrix eliminates the need for repetitive design rules and accurately classifies X-ray images of RA patients into five categories representing different stages of joint erosion and assessing joint damage. This innovative approach improves the diagnosis of RA and facilitates timely and effective treatment. To improve the RA classification task, we introduce the *ShuffleAttention (SA)* module and integrate it with a multilayer residual block, known as the *ResSA block*. This combination improves the ability of the network to extract spatial features at different scales. In addition, we replace short-cut links with dense residual links, referred to as *densepath*, to address the problem of misalignment between encoder and decoder features. These modifications significantly improve the network's performance in RA classification compared to state-of-the-art approaches. By accurately and efficiently diagnosing RA, MediMatrix can contribute to improved patient care and management. The symbols used in this paper are summarized in Table 1. The main contributions of this paper are as follows:

- We present *MediMatrix*, an innovative network for classifying medical images that automates, speeds up, and standardizes the assessment of affected areas in patients with rheumatoid arthritis. This innovative tool has the potential to transform the diagnosis and management of rheumatoid arthritis by enabling more precise and efficient care.
- To improve our network's multi-scale spatial feature extraction capability, we introduce the SA module. This module is integrated at the end of the convolutional block. Our network can dynamically adjust the channel attention weights of feature maps by combining the SA module with the multiple residual structure (ResSA block) and replacing short-cut connections with dense residual connections between comparable levels (densepath). As a result, our network is able to represent the features better and discriminate similarities more accurately across different categories.
- To pre-train models for natural image classification, we used grayscale ImageNet with MediMatrix instead of traditional methods on color ImageNet. Our findings suggest that color is not necessary for effective natural image classification since grayscale models performed comparably to color models on the original ImageNet challenge. Through an extensive experimentation process, we discovered that gray-

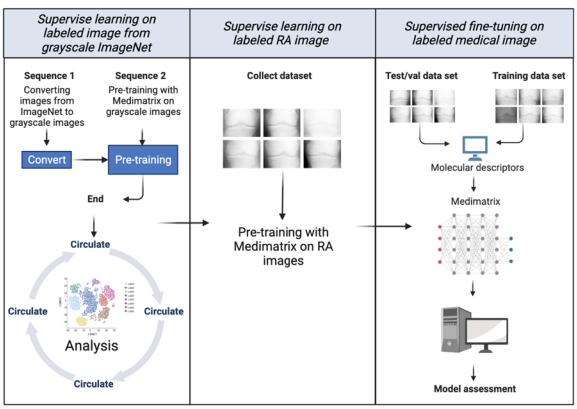

Fig. 1 Overall flow chart of the method we use

Table 1 Definition of symbols

Symbol	Meaning
F^c_{avg}	Average pooling in the channel attention
F^c_{max}	Max pooling in the channel attention
F^s_{avg}	Average pooling in the spatial attention
F^s_{max}	Max pooling in the spatial attention
$\mu F1$	The harmonic mean of Precision and Recall
K	The number of categories
I	The indicator function
p_k	The probability that the category is k
N	The total number of samples

scale models, on average, outperformed color models, improving accuracy by $(1.32 \pm 0.1)\%$.

Related work

We present a comprehensive classification of RA using both traditional and deep learning methods.

RA image classification based on traditional methods

This section aims to provide a detailed introduction to the application of traditional methods in RA image classification, with a focus on risk assessment and early detection.

Assessing the risk of developing RA

O'Neillet et al. [13] proposed a serum proteomic-based regression model to assess arthritis risk. However, this model's limitation is its poor prediction performance for ACPA-positive individuals. Ou et al. [6] utilized frequency domain phase spectrum to quantify joint space narrowing progression in finger joint images at baseline and follow-up. But their approach was labor-intensive.

Using holographic data in the diagnosis of RA

Computer-aided diagnostic methods have shown comparable accuracy to dynamic contrast-enhanced MRI, with the advantage of reduced image analysis time. Nevertheless, there are limited experimental cases using ultrasound images for synovitis classification and quantification [14, 15]. Alarcon-Paredes et al. [16] used random forest and wrapper feature selection, but encountered significant computational overhead. Aizenberg et al. [17] used atlas-based segmentation and fuzzy C-means clustering, but encountered problems with noisy data and dependence on initial values, limiting generalization across datasets.

Using clinical and sensor data to diagnose RA

Fukae et al. [18] implemented an approach that converts clinical data into two-dimensional array images and uses CNN to categorize rheumatoid arthritis (RA) patients. The algorithm produced results that were consistent with the diagnoses of three rheumatology experts. However, the sensitivity of the algorithm was significantly affected by the size of the input images. On the other hand, Bardhan et al. [19] developed a two-stage classification algorithm capable of accurately labeling about three-quarters of knee thermal imager scans. The first stage detects knee joints affected by arthritis, while the second stage identifies knee joints affected by RA.

RA image classification based on deep learning

The application of deep learning methods in assessing the risk of RA has gradually increased.

Using holographic data in the diagnosis of RA

Chocholova et al. [20] used glycomic techniques and serum samples to differentiate healthy individuals, serum-positive RA patients, and serum-negative RA patients. They combined anti-CCP and total RF measurements with RCA carbohydrate analysis based on ELLBA, achieving high accuracy. However, the method's complexity hinders widespread adoption. Heard et al. [21] utilized artificial neural network and decision tree methods to classify healthy individuals, RA patients, and OA patients based on a panel of inflammatory cytokines from serum samples. However, this method demands extensive training data and computational resources, leading to long training times and overfitting susceptibility, and it does not address image noise influence.

Using clinical and sensor data in the diagnosis of RA

Fukae et al. [18] used AlexNet and ResNet-18 to convert clinical data into two-dimensional array images for arthritis diagnosis. However, this method requires significant computational resources and incurs high costs. It is also sensitive to the size of the input image and prone to overfitting. Wyns et al. [22] used the Kochnin neural network (including self-organizing maps) to predict the diagnosis of early arthritis patients. However, the experimental samples did not include indeterminate samples.

Using imaging data to diagnose RA

Wu and colleagues used DenseNet to classify synovial hyperplasia in ultrasound images to assess RA severity [15]. However, this method is prone to underfitting and overfitting, resulting in reduced generalizability, and it requires significant memory resources. Hirano et al. [23] used CNN to evaluate imaging findings of joint destruction in rheumatoid arthritis, but the accuracy of the method is extremely low. Murakami et al. [15] used the MSGVF snake algorithm and DCNN classifier to identify osteoporosis. They used a triple cross-validation method to validate independent test datasets. However, the training process of this method requires a amount of memory resources.

Discussion

Comparing with the above methods, MediMatrix has the following advantages: (1) it innovatively employs grayscale images for pre-training, improving the RA

detection accuracy; (2) it includes an innovative attention mechanism, which removes the necessity of specifying the size of the input image and helps the network extract features in multi-scale space; and (3) it utilizes short-circuit linking to dynamically regulate the attention weights of the channels of the feature maps, thus improving feature representation.

MediMatrix

Overall network architecture

The MediMatrix network architecture (Fig. 2) is built based on inspiration from a variant of U-Net [12] to enhance its performance. The network can be divided into three main parts: pre-training, the feature encoding and decoding part.

During the pre-training phase, we use the proposed MediMatrix in an innovative way to train the grayscale ImageNet. Then, we utilize the weights obtained from the pre-training to initialize the same model for labeled RA images. Lastly, we fine-tune the obtained weights using our proposed model.

In the coding part, we replace the convolution block in the traditional U-net with an improved Multi-ResNet block, called ResSA block, proposed in "ResSA block" section. This modification helps to capture more comprehensive and informative features.

In the decoding section, an innovative approach called Densepath is introduced in "Densepath" section. The purpose of this technique is to overcome the semantic gap that exists between the encoder and decoder features. Unlike the traditional U-Net architecture, which uses a direct connection between the encoder and decoder, Densepath allows for an improved flow of information, resulting in more accurate reconstructions. This method enhances the network's ability to generate accurate representations of the input data, resulting in improved diagnostic accuracy for arthritis damage detection using X-ray images.

Given that the task involves medical image classification, we include a softmax layer at the end of the decoding part to generate prediction probabilities for the input image across five categories. All convolutional blocks in the network utilize the Rectified Linear Unit (ReLU) activation function and are normalized.

ResSA block

The ResSA block (Fig. 3), represents a sophisticated approach to image feature extraction. This advanced block uses three sets of 3×3 convolutional blocks, each with dense connections for optimal efficiency. In addition, to better incorporate spatial information into the model, a 1×1 convolutional block is used for the residual

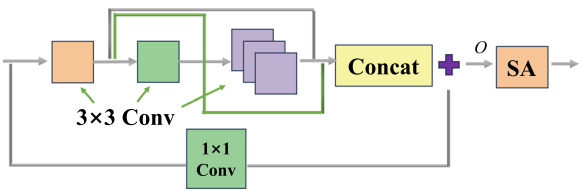

Fig. 3 The structure of ResSA Block

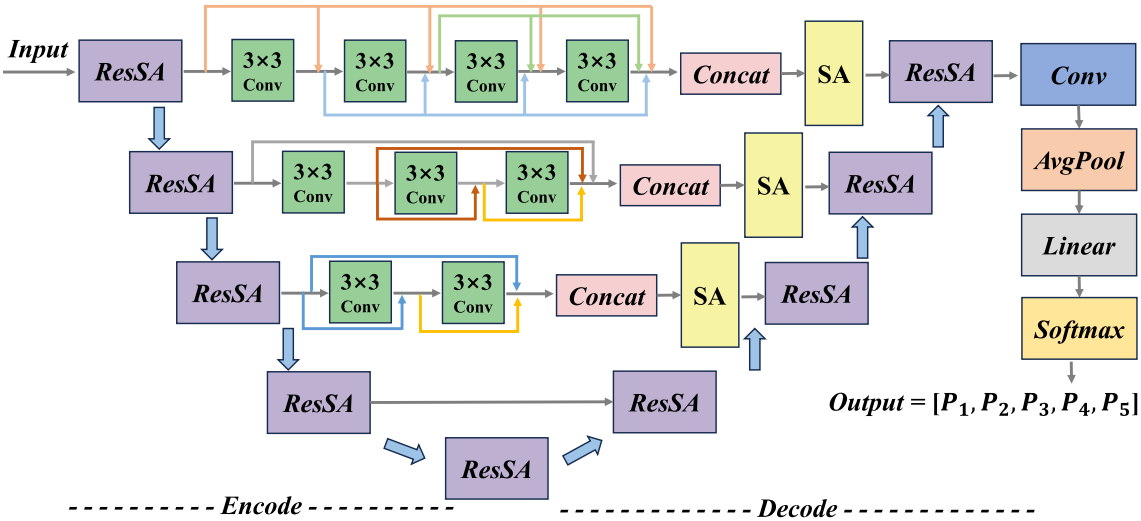

Fig. 2 Detailed structure of MediMatrix

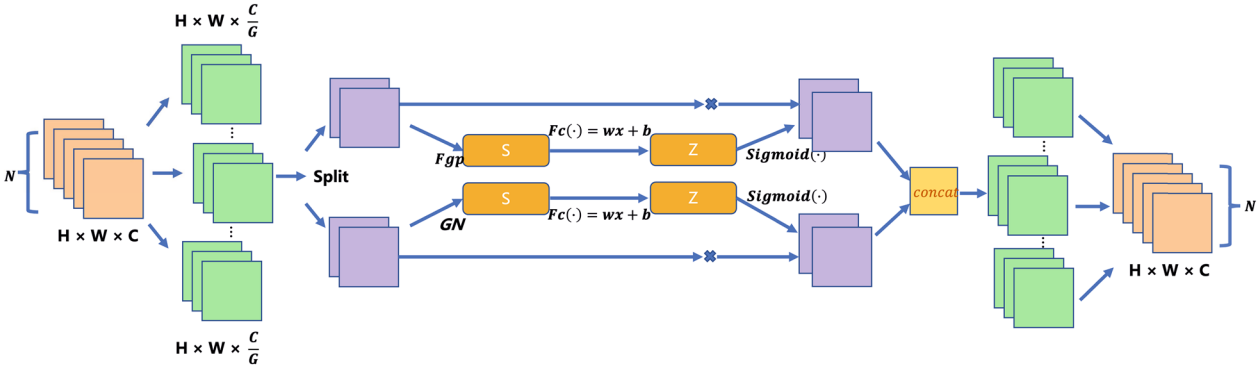

Fig. 4 General structure of SA

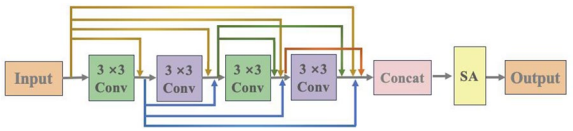

Fig. 5 Detailed structure of Densepath

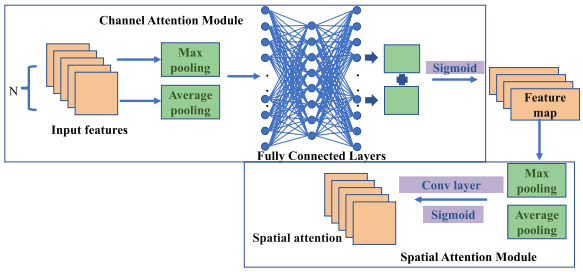

Fig. 6 The process of feature processing in SA

connection. By summing the outputs of both the dense and the residual connections, the resulting output of the ResSA block (denoted *O* here) emerges as a powerful representation of the original image.

To enhance the convolution layer's ability to select and integrate multi-channel subfeatures, we introduced the SA module (Fig. 4). This module divides the combined output *O* of the convolutional layer into multiple sets of subfeatures. Each subfeature is individually processed using spatial and interchannel attention mechanisms, resulting in processed subfeatures. The processed subfeatures are then merged via the channel shuffle operation to fuse the features. This approach enhances the model's selective emphasis on the most relevant information and facilitates the integration of diverse features, thereby improving performance.

Densepath

The Densepath (Fig. 5), addresses the problem of incompatible feature fusion in the U-Net decoding process. The traditional U-Net architecture uses shortcut connections between corresponding layers, which can cause low-level feature information to propagate to the high-level decoding network, leading to feature fusion incompatibility.

To solve this problem, we propose to replace the shortcut connection with a dense residual connection. This approach ensures a more balanced propagation of feature information throughout the network and improves feature fusion compatibility. By incorporating the dense residual connection, the densepath bridges the semantic gap between encoder and decoder features, enabling accurate feature reconstruction.

ShuffleAttention

SA (Alg. 1) is an efficient attention mechanism module based on the partial structure of the Convolutional Block Attention Module (CBAM) [24], which combines spatial and channel information, in contrast to Senet, which focuses only on the channel. The feature processing is shown in Fig. 6.

The channel attention module works in the following steps: First, the input feature map undergoes global max pooling and global average pooling separately based on its width and height. Next, the resulting outputs are fed into a fully connected layer. The element-wise multiplication of this Multi-Layer

Perceptron (MLP) output and the input feature map generates the necessary input features for the Spatial Attention. These steps form the channel attention mechanism.

The channel attention module performs spatial dimension compression on the feature map by obtaining a one-dimensional vector and applying operations to it. This compression includes both average pooling and maximum pooling to aggregate spatial information from the feature map. The aggregated information is then passed through a shared network to compress the spatial dimensions of the input feature map. The resulting compressed feature map is summed and fused element-wise to produce a channel attention map. Channel attention focuses on the importance of each element in the feature map. Mean pooling provides feedback for every pixel point in the feature map, while max pooling provides feedback for gradients only at locations where the response is highest in the feature map during gradient back-propagation calculations (Eq. 1).

$$\begin{aligned}\mathbf{M_c}(\mathbf{F}) &= \sigma(\mathrm{MLP}(\mathrm{AvgPool}(\mathbf{F})) + \mathrm{MLP}(\mathrm{MaxPool}(\mathbf{F}))) \\ &= \sigma\left(\mathbf{W_1}\left(\mathbf{W_0}\left(\mathbf{F^c_{avg}}\right)\right) + \mathbf{W_1}\left(\mathbf{W_0}\left(\mathbf{F^c_{\max}}\right)\right)\right)\end{aligned} \tag{1}$$

where σ is a sigmoid operation, *MLP* is a multilayer perceptron, *F* stands for feature. F^c_{avg} and F^c_{max} are the average and maximum pooling in the channel attention module.

The *MaxPool* operation extracts the maximum value across the channel for each spatial location, resulting in a feature map with the same height and width. The *AvgPool* operation calculates the average value across the channel for each spatial location, resulting in a feature map with the same height and width. These two extracted feature maps, each with a single channel, are then combined to obtain a new feature map (Eq. 2).

$$\begin{aligned}\mathbf{M_s}(\mathbf{F}) &= \sigma\left(f^{7\times7}([\mathrm{AvgPool}(\mathbf{F}); \mathrm{MaxPool}(\mathbf{F})])\right) \\ &= \sigma\left(f^{7\times7}\left(\left[\mathbf{F^s_{avg}}; \mathbf{F^s_{\max}}\right]\right)\right)\end{aligned} \tag{2}$$

where F^s_{avg} and F^s_{max} are the average pooling and maximum pooling in the spatial attention module. 7×7 indicates the size of the convolution kernel, a 7×7 convolution kernel works better than a 3×3 convolution kernel.

Algorithm 1: ShuffleAttention

Input: input_tensor: input tensor
Output: Output tensor with attention

function ChannelAttention(*inputs*)

1. avg_pool ← GlobalAveragePooling2D(inputs)
2. max_pool ← GlobalMaxPooling2D(inputs)
3. fc1 ← Dense(units = num_filters/8, activation = ReLU)(avg_pool)
4. fc2 ← Dense(units = num_filters, activation = ReLU)(fc1)
5. channel_attention ← Multiply($[fc2, \text{inputs}]$)

return channel_attention

function SpatialAttention(*inputs*)

1. conv1 ← Conv2D(filters = 1, kernel_size = $(3, 3)$, padding = same)(inputs)
2. sigmoid ← Activation(sigmoid)(conv1)
3. spatial_attention ← Multiply($[sigmoid, \text{inputs}]$)

return spatial_attention

function ShuffleAttention(*inputs*)

1. channel_attention ← ChannelAttention(inputs)
2. spatial_attention ← SpatialAttention(channel_attention)

return spatial_attention

Loss function

MediMatrix uses the Eq. (3) as loss function. This function quantifies the dissimilarity between two probability distributions: the actual probability distribution and the predicted probability distribution of the same random variable. By evaluating the cross-entropy loss, the model can assess how well it approximates the target classes by comparing the predicted probabilities to the actual probabilities. A lower value of the cross-entropy loss indicates a higher model performance in terms of accurately predicting the target classes.

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} I\left(y_i = k\right)\log\left(p_k\right) \tag{3}$$

$$I(x) = \begin{cases} x = 1 & \text{True} \\ x = 0 & \text{False} \end{cases} \quad (4)$$

where N is the total number of samples, K is the number of categories, I is the indicator function, and p_k represents the probability that the category is k

Experiment

Datasets

The article presents representative images of each category in the knee osteoarthritis dataset (Fig. 7). The dataset comprises a total of 9786 samples. For assessing the severity of osteoarthritis in the knee radiographs, the Kellgren–Lawrence (K–L) classification, a qualitative assessment method, was employed. Specially trained medical raters graded the arthritis severity on the radiographs, providing labeled training data for MediMatrix. The different disease stages represented by the grading are outlined in Table 2.

For data partitioning, we divide the data set into a training set (70%), a validation set (20%), and a test set (10%). This division allows us to train and optimize MediMatrix on the training set, tune the hyperparameters using the validation set, and evaluate the model's performance on the unseen test set. To ensure equal sensitivity to each category, the number of images for each category is consistent across all datasets.

We used the Knee X-ray Images dataset [25] for scalability experiments (Table 3). The dataset contains DICOM-standard images with dimensions of 1345 × 2455. The images were collected based on demographic features, including age, gender, blood type, occupation, and weight. Among the 532 patients, there were 301 females and 231 males.

Fig. 7 Selected samples from the dataset

Table 3 KL grades assigned by 2 Medical Experts

KL grade	Medical Expert-I	Medical Expert-II
Normal (G-0)	337	348
Doubtful (G-1)	139	128
Mild (G-2)	31	31
Moderate (G-3)	9	9
Severe (G-4)	16	16

Experiment settings and metrics

In our experimental evaluation of the multi-label classification of knee joint severity, we used the following evaluation indicators: $\mu F1$, balanced accuracy, AUC, and Cohen's Kappa score.

Micro $F1$ ($\mu F1$): The $F1$ score is the harmonic mean of Precision and Recall. $\mu F1$ is the average of the $F1$ scores calculated for all categories, regardless of class imbalance. We utilize the $F1$ score to assess the overall multi-label classification capability of the model.

Balanced Accuracy (BA): Balanced accuracy is the average accuracy calculated for all categories, taking into account the sample imbalance between different categories. It is obtained by assigning weights to the accuracy of each category.

Area Under the Curve (AUC): AUC is the area under the receiver operating characteristic (ROC) curve. The ROC curve represents the relationship between the true positive rate (recall rate) and the false positive rate. AUC measures the performance of the classifier at different thresholds and is an effective measure of accuracy.

Cohen's Kappa Score: Cohen's Kappa is a statistical index used to measure the consistency between the classifier and the random selection. It takes into account the correctness of the classification and the influence of random selection. MediMatrix is a multi-class imbalance problem, and Cohen 's Kappa can well evaluate the multi-class classification ability.

Using these metrics, we can evaluate the performance of the classification model in terms of multi-label knee

Table 2 Categories of datasets and their detailed explanation

Stage	Description
0	Healthy knee joint
1	Joint space suspected narrowing, there may be osteophytes
2	Obvious osteophyte, suspicious narrowing of joint space
3	Moderate osteophytes, joint space narrowing more obvious, there are hardening changes
4	A large number of osteophytes, joint space significantly narrowed, severe sclerosing lesions and obvious deformity

Table 4 Results of grayscale and color images on SimCLR

Architecture	Color		Grayscale	
	Top-1 (%)	Top-5 (%)	Top-1	Top-5
ResNet-50 (1×)	68.33	88.02	67.92	87.64
ResNet-50 (2×)	73.24	91.40	72.91	90.92
ResNet-50 (4×)	76.35	93.12	75.98	92.85

joint severity. In addition, given the similarity in severity between grades 1, 2 and 3, we also calculated the AUC (one-to-one) between grades 1 and 2 and between grades 2 and 3 to specifically measure the classification performance in discriminating between these pairs.

Our model is built in a pytorch box μ frame. All samples are standardized to facilitate subsequent computation and storage. The epoch in the model is set to 100, the AdamW with a learning rate of 1e-3 is updated, and the experiment runs on an NVIDIA GeForce GTX 3090 GPU and an Intel Core i7-12700 H CPU.

Pretraining protocol

This study investigated the effectiveness of pre-training using MediMatrix. No significant difference was observed in the classification performance of grayscale and color images. Additionally, this was demonstrated using SimCLR [26]—ResNet-50 (1×), ResNet-50 (2×), and ResNet-50 (4×) [27]. Complying with the approach proposed by SimCLR [26], two fully connected layers were used to transform the ResNet outputs into 128-dimensional embeddings (Table 4).

To assess the performance of ImageNet models, we measured Top-1 and Top-5 accuracy for various architectures. Our aim was to compare the effectiveness of color and grayscale images, demonstrating that color is not the primary criterion for ImageNet image classification. Surprisingly, the model trained and evaluated on grayscale ImageNet performed just $(0.4 \pm 0.07)\%$ worse than the color model. Moreover, our results emphasize that pre-training on ImageNet complements pre-training on unlabeled medical images, highlighting its importance in improving performance. We further validated our findings by experimenting with other state-of-the-art (SOTA) methods (Table 5).

Table 6 Experimental performance after fine-tuning

		Color	Grayscale
Method	Dataset	Acc (%)	Acc (%)
MediMatrix	Kaggle	90.14	91.36
	Knee X-ray Images	87.32	88.72

Fine-tuning protocol

During the fine-tuning process, we initialize the weights of the pre-trained network with the aim of utilizing them for the downstream task.

To optimize the performance of each combination of pretraining approach and downstream fine-tuning task, we conduct a comprehensive hyperparameter search. This involves performing a grid search across seven logarithmically spaced learning rates, ranging from $10^{-3.5}$ to $10^{-0.5}$, as well as three logarithmically spread weight decay values, ranging from 10^{-5} to 10^{-3}. Through this search process, we determine the optimal learning rate and weight decay for each specific case.

We apply the same search approach when training from the supervised training baseline. Remarkably, we find that regardless of the fine-tuning settings, achieving optimal performance typically requires 100 epochs of training. In addition, based on extensive experiments, our grayscale model outperforms the color model in improving the disease recognition rate (Table 6). Our grayscale model improves the average accuracy of disease by about $(1.32 \pm 0.1)\%$.

Table 5 Classification results on grayscale and color images on top of ImageNet

Method	Architecture	Color		Grayscale	
		Top-1 (%)	Top-5 (%)	Top-1 (%)	Top-5 (%)
MoCo [28]	ResNet-50 (1×)	60.63	–	60.12	–
PIRL [29]	ResNet-50 (1×)	63.54	–	63.24	–
CPCv2 [30]	ResNet-50 (1×)	63.75	85.37	63.28	84.74
CPCv2 [30]	ResNet-161 (∗)	71.53	90.01	70.98	89.65
MoCo [28]	ResNet-50 (4×)	60.45	–	60.03	–
CMC [31]	ResNet-50 (2×)	68.49	88.25	67.98	87.83
BigBiGAN [32]	RevNet-50 (4×)	61.24	81.92	60.83	81.57

Comparative experiments

In order to assess the effectiveness of our preprocessed dataset, we conducted ample comparative experiments with five different methods on the preprocessed dataset: MobileNetv2 [33], CNN + Ordered Loss [34], the Extrusion-Excitation Block (SE Block) [35], DeepKnee [36], and Set [37].

Designed specifically for image classification and object detection tasks on devices with limited computing resources, MobileNetv2 uses lightweight design and deep separable convolution with linear bottleneck structures to reduce computation and model size. CNN + Ordinal Loss addresses overfitting in image classification by incorporating ordinal loss, a loss function used to handle ordered categories, into the CNN classification model to improve its performance. The Squeeze-Excitation Block (SE Block) [38] is a mechanism that improves the representational ability of convolutional neural networks by adaptively adjusting the importance of each channel in the feature map. It includes a squeeze phase and an excitation phase, learning the relationship between feature channels to improve the model's representational ability. DeepKnee is a neural network specifically designed for automatic analysis and diagnosis of knee X-ray images, using convolutional neural networks to extract features, classify and predict joint disease. Finally, Ensemble, which combines the prediction results of multiple basic models, enhances classification performance by using models of different architectures or initialized with different training data and parameters, leading to improved accuracy in the prediction.

Quantitative results: The quantitative results (Table 7) of our comparative experiments demonstrate the superiority of MediMatrix over the five evaluated methods in terms of $\mu F1$, balance accuracy, and Cohen's Kappa score. MediMatrix achieved the best performance in these evaluation metrics, indicating its effectiveness in accurately classifying the degree of arthritis damage.

In terms of $\mu F1$, MediMatrix outperformed all other methods, achieving the highest average $F1$ score across all categories. This indicates that our model has a strong multi-label classification ability and can effectively capture the nuances of different knee joint severity levels.

Furthermore, MediMatrix exhibits the highest balance accuracy, effectively considering the sample imbalance among different categories. This achievement demonstrates our model's superior ability to achieve balanced and accurate classification performance across all severity levels, even in the presence of variations in sample distribution.

A significant achievement is that the AUC results of MediMatrix are comparable to those of the SE block. Our model demonstrated competitive performance in distinguishing between specific severity levels (e.g., grade 1 and 2) with AUC being a widely used metric to measure classifier performance under different thresholds. These results indicate MediMatrix's effectiveness in discriminating intermediate severity levels, which are often challenging to classify accurately.

Furthermore, our method achieved the highest Cohen's Kappa value of 0.6846 among all methods. Cohen's Kappa measures the agreement between classifier predictions and pre-annotated scores, taking into account the effect of random selection. The high Cohen's Kappa value indicates a strong agreement between our model's predictions and the ground truth scores assigned by medical raters, implying a higher level of reliability and accuracy in our classification results.

Overall, these quantitative results demonstrate the superior performance of MediMatrix compared to the evaluated methods, underscoring its efficacy in accurately classifying knee joint severity and predicting consistency with pre-annotated scores. The visualized data for comparison with the results of some SOTA methods are shown in Fig. 8. As shown in it, our proposed model has an overall advantage over other SOTA methods.

Ablation experimentnt

In our ablation experiments, we examined the effectiveness of two key components in MediMatrix: the attention mechanism (SA) and the respath module. To evaluate the attention mechanism, we removed the SA module from the convolution block in the base

Table 7 Comparison of results with some SOTA methods (Bold: The best, Italics: The second best)

Method	$\mu F1$	BA	AUC	Kappa	AUC(1,2)	AUC(2,3)
MobileNetV2	0.5104	0.3532	0.7822	0.2554	0.6208	0.6191
CNN + Ordinal loss	0.6865	0.6638	0.8950	0.5557	0.7298	0.8576
SE block	0.7336	0.7237	0.9237	0.6237	0.7866	0.9265
DeepKnee	0.3956	0.5078	0.7456	0.2287	0.5931	0.7398
Ensemble	0.7405	0.7342	0.7342	0.6327	0.7896	0.9360
MediMatrix	**0.7941**	**0.8059**	*0.9136*	**0.6846**	**0.8948**	0.8865

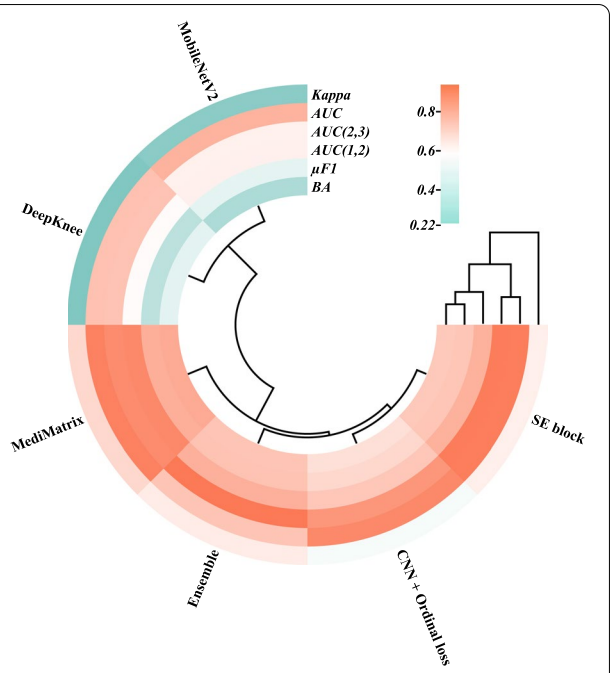

Fig. 8 Visualization results of comparative experiments

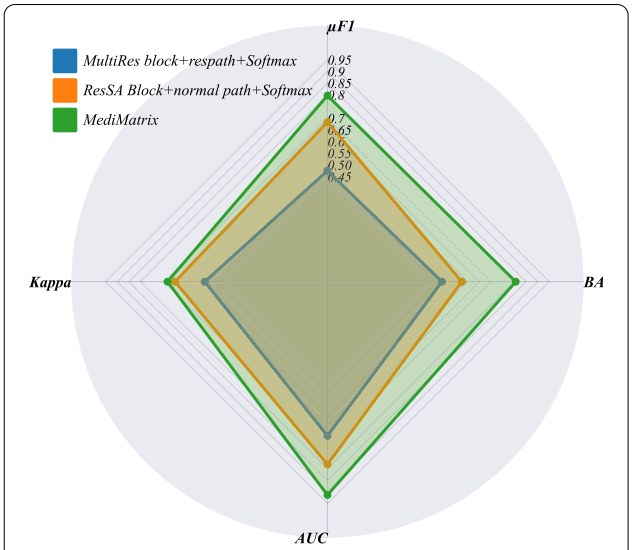

Fig. 9 Visualization results of ablation experiments

MultiResUNet architecture. The network architecture for this attention ablation experiment consisted of the MultiRes block, Respath, and Softmax layers. The results showed a decrease in network accuracy and Cohen's Kappa value (0.5257), indicating the critical role of the SA module in improving spatial feature extraction at different scales. The attention mechanism in MediMatrix enhances the extraction and integration of multi-scale features, resulting in improved classification performance.

Next, we investigated the effectiveness of the Respath module, which compensates for spatial information loss during encoder-to-decoder propagation. In this ablation experiment, we replaced the respath module with a direct link. The network architecture for this Respath ablation experiment included the ResSA block, normal path (without Respath), and Softmax layers (Fig. 9).

The inclusion of the Respath module allowed the network to retain more spatial information during encoder-to-decoder propagation. As a result, the accuracy of the network showed a significant improvement over the Respath-enabled configuration. The specific performance metrics are shown in Fig. 9. These results highlight the importance of the Respath module in preserving spatial information and its impact on improving the overall accuracy of MediMatrix.

In summary, the ablation experiments confirmed the effectiveness of the attention mechanism and the Respath module in MediMatrix. The ShuffleAttention module improved multiscale spatial feature extraction, resulting in improved accuracy and Cohen's Kappa Score. The respath module compensated for spatial information loss and significantly improved network accuracy. These modules played a critical role in achieving the superior performance of MediMatrix, as demonstrated by the results.

Generalizability experiments

To assess the generalizability of the MediMatrix classification framework, we performed generality proof experiments using the Knee X-ray Images dataset [25]. This dataset consists of X-ray images, similar to our original dataset, and contains five categories. The dataset was divided into training, test and validation sets in a ratio of 7:2:1. The results of MediMatrix on the training, validation, and test sets are 91.45%, 89.23% and 88.72%.

Case study

MediMatrix holds great promise for future RA treatment and integration into intelligent healthcare facilities. This method can improve the diagnostic accuracy and decision-making capabilities of medical professionals who can use it to optimise the treatment of RA patients. It can streamline diagnosis, reduce the time required to make an accurate diagnosis, and initiate timely treatment. By using a data-driven approach to RA diagnosis, MediMatrix holds the promise of improving patient outcomes. In addition, by seamlessly integrating with electronic health records, patient monitoring systems, and telemedicine platforms, MediMatrix has the potential to facilitate interdisciplinary collaboration among healthcare

professionals. It can also contribute to a more holistic approach to healthcare.

Conclusion

The paper introduces MediMatrix, a novel network that uses pre-trained grayscale images for assessing thermographic images in patients with RA in a fast and automated manner. Shuffle attention has been added to MediMatrix to improve multiscale spatial feature extraction by replacing shortcut connections with dense residual connections, which reduces parallax. The experimental results demonstrate that MediMatrix is superior to the existing methods and has achieved higher diagnostic accuracy on the RA X-ray dataset. Through the use of deep learning and attention mechanisms, MediMatrix provides a cost-effective diagnostic approach and a reliable automated solution for RA diagnosis. This benefits healthcare professionals and patients through improved medical decision-making and care. In the future, we will examine how self-supervised learning can enhance MediMatrix by addressing the problem of inadequately labeling medical images.

Data availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Declarations

Conflict of interest

We affirm that we have no commercial or associative interests that could create a conflict of interest related to the submitted work.

Author details

[1]Department of Rheumatology, Zhongda Hospital, School of Medicine, Southeast University, Nanjing 210009, China. [2]Department of Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science and Technology, Nanjing 210044, China.

Received: 31 July 2023 Accepted: 8 September 2023
Published: 26 September 2023

References

1. Goebel A, et al. The autoimmune aetiology of unexplained chronic pain. Autoimmun Rev. 2022;21:103015.
2. Nakatsu K, Morita K, Yagi N, Kobashi S. Finger joint detection method in hand x-ray radiograph images using statistical shape model and support vector machine, 2020, pp. 1–5. IEEE.
3. Ainsworth RI, et al. Systems-biology analysis of rheumatoid arthritis fibroblast-like synoviocytes implicates cell line-specific transcription factor function. Nat Commun. 2022;13:6221.
4. Xie Y, Richmond D. Pre-training on grayscale imagenet improves medical image classification; 2018.
5. Tschandl P, Rosendahl C, Kittler H. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data. 2018;5:1–9.
6. Ou Y, et al. A sub-pixel accurate quantification of joint space narrowing progression in rheumatoid arthritis. IEEE J Biomed Health Inform. 2022;27:53–64.
7. Zhou Q, Huang Z, Ding M, Zhang X. Medical image classification using light-weight CNN with spiking cortical model based attention module. IEEE J Biomed Health Inform. 2023;27:1991–2002.
8. Qu Z, Sun H. A secure information transmission protocol for healthcare cyber based on quantum image expansion and grover search algorithm. IEEE Trans Netw Sci Eng. 2022.
9. Mate GS, Kureshi AK, Singh BK. An efficient CNN for hand x-ray classification of rheumatoid arthritis. J Healthc Eng. 2021.
10. Rohrbach J, Reinhard T, Sick B, Dürr O. Bone erosion scoring for rheumatoid arthritis with deep convolutional neural networks. Comput Electr Eng. 2019;78:472–81.
11. Tan W, et al. Segmentation of lung airways based on deep learning methods. IET Image Proc. 2022;16:1444–56.
12. Ibtehaz N, Rahman MS. Multiresunet: rethinking the u-net architecture for multimodal biomedical image segmentation. Neural Netw. 2020;121:74–87.
13. O'Neil LJ, et al. Proteomic approaches to defining remission and the risk of relapse in rheumatoid arthritis. Front Immunol. 2021;12:729681.
14. Hu X, et al. Joint landmark and structure learning for automatic evaluation of developmental dysplasia of the hip. IEEE J Biomed Health Inform. 2021;26:345–58.
15. Wu M, et al. A deep learning classification of metacarpophalangeal joints synovial proliferation in rheumatoid arthritis by ultrasound images. J Clin Ultrasound. 2022;50:296–301.
16. Alarcón-Paredes A, et al. Computer-aided diagnosis based on hand thermal, RGB images, and grip force using artificial intelligence as screening tool for rheumatoid arthritis in women. Med Biol Eng Comput. 2021;59:287–300.
17. Aizenberg E, et al. Automatic quantification of bone marrow edema on MRI of the wrist in patients with early arthritis: a feasibility study. Magn Reson Med. 2018;79:1127–34.
18. Fukae J, et al. Convolutional neural network for classification of two-dimensional array images generated from clinical information may support diagnosis of rheumatoid arthritis. Sci Rep. 2020;10:1–7.
19. Bardhan S, Bhowmik MK. 2-stage classification of knee joint thermograms for rheumatoid arthritis prediction in subclinical inflammation. Australasian Phys Eng Sci Med. 2019;42:259–77.
20. Chocholova E, et al. Glycomics meets artificial intelligence-potential of glycan analysis for identification of seropositive and seronegative rheumatoid arthritis patients revealed. Clin Chim Acta. 2018;481:49–55.
21. Heard BJ, et al. A computational method to differentiate normal individuals, osteoarthritis and rheumatoid arthritis patients using serum biomarkers. J R Soc Interface. 2014;11:20140428.
22. Wyns B, et al. Prediction of diagnosis in patients with early arthritis using a combined Kohonen mapping and instance-based evaluation criterion. Artif Intell Med. 2004;31:45–55.
23. Hirano T, et al. Development and validation of a deep-learning model for scoring of radiographic finger joint destruction in rheumatoid arthritis. Rheumatol Adv Pract. 2019;3:rkz047.
24. Woo S, Park J, Lee, J-Y, Kweon IS. Cbam: convolutional block attention module, 2018; pp. 3–19.
25. Gornale S, Patravali P. Digital knee x-ray images. Mendeley Data 2020;**1**.
26. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations, pp. 1597–1607 (PMLR, 2020).
27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition, 2016; pp. 770–778.
28. He K, Fan H, Wu Y, Xie S, Girshick, R. Momentum contrast for unsupervised visual representation learning, 2020; pp. 9729–9738.
29. Cao Z, Yu H, Yang H. Sano A. Pirl: participant-invariant representation learning for healthcare using maximum mean discrepancy and triplet loss. arXiv preprint arXiv:2302.09126 2023.
30. Henaff O. Data-efficient image recognition with contrastive predictive coding, pp. 4182–4192 (PMLR, 2020).
31. Tian Y, Krishnan D, Isola P. Contrastive multiview coding. New York: Springer; 2020. p. 776–94.
32. Donahue J, Simonyan K. Large scale adversarial representation learning. Adv Neural Inf Process Syst 2019;**32**.

33. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition 2018.
34. Chen P, Gao L, Shi X, Allen K, Yang L. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. Comput Med Imaging Graph. 2019;75:84–92.
35. Jie et al. Squeeze-and-excitation networks. IEEE Trans Pattern Anal Machine Intell 2019.
36. Tiulpin A, Thevenot J, Rahtu E, Lehenkari P., Saarakkala, S. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach; 2017.
37. Tiulpin A, Saarakkala S. Automatic grading of individual knee osteoarthritis features in plain radiographs using deep convolutional neural networks. Diagnostics. 2020;10:932.
38. Hu J, Shen L, Sun G. Squeeze-and-excitation networks, 2018; pp. 7132–7141.