

RESEARCH

Open Access



Machine learning techniques for computer-aided classification of active inflammatory sacroiliitis in magnetic resonance imaging

Matheus Calil Faleiros¹, Marcello Henrique Nogueira-Barbosa^{2,3,4,5*} , Vitor Faeda Dalto⁴, José Raniery Ferreira Júnior^{2,3}, Ariane Priscilla Magalhães Tenório², Rodrigo Luppino-Assad², Paulo Louzada-Junior², Rangaraj Mandayam Rangayyan⁶ and Paulo Mazzoncini de Azevedo-Marques^{2,3}

Abstract

Background: Currently, magnetic resonance imaging (MRI) is used to evaluate active inflammatory sacroiliitis related to axial spondyloarthritis (axSpA). The qualitative and semiquantitative diagnosis performed by expert radiologists and rheumatologists remains subject to significant intrapersonal and interpersonal variation. This encouraged us to use machine-learning methods for this task.

Methods: In this retrospective study including 56 sacroiliac joint MRI exams, 24 patients had positive and 32 had negative findings for inflammatory sacroiliitis according to the ASAS group criteria. The dataset was randomly split with ~80% (46 samples, 20 positive and 26 negative) as training and ~20% as external test (10 samples, 4 positive and 6 negative). After manual segmentation of the images by a musculoskeletal radiologist, multiple features were extracted. The classifiers used were the Support Vector Machine, the Multilayer Perceptron (MLP), and the Instance-Based Algorithm, combined with the Relief and Wrapper methods for feature selection.

Results: Based on 10-fold cross-validation using the training dataset, the MLP classifier obtained the best performance with sensitivity = 100%, specificity = 95.6% and accuracy = 84.7%, using 6 features selected by the Wrapper method. Using the test dataset (external validation) the same MLP classifier obtained sensitivity = 100%, specificity = 66.7% and accuracy = 80%.

Conclusions: Our results show the potential of machine learning methods to identify SIJ subchondral bone marrow edema in axSpA patients and are promising to aid in the detection of active inflammatory sacroiliitis on MRI STIR sequences. Multilayer Perceptron (MLP) achieved the best results.

Keywords: Magnetic resonance imaging, Sacroiliac joint inflammation, Spondyloarthritis, Machine learning, Artificial intelligence, Computer-assisted diagnosis

* Correspondence: marcello@fmrp.usp.br

²Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil

³MAInLab Medical Artificial Intelligence Laboratory, Ribeirão Preto Medical School, Ribeirão Preto, Brazil

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

The term spondyloarthritis (SpA) encompasses a group of diseases characterized by inflammation in the spine and in the peripheral joints, as well as other clinical features. The current concept of the spectrum of SpA comprises axial spondyloarthritis (axSpA) and peripheral spondyloarthritis. In recent years, there has been tremendous progress in understanding the natural history and pathogenetic mechanisms underlying SpA, leading to the development of effective treatments. It has become imperative to identify the disease early and accurately, to offer patients effective treatment in a safe manner [1]. SpA usually starts in the young adult age. Its progression frequently contributes to significant physical disability and decreased quality of life if early diagnosis and early treatment are not achieved. This group of diseases presents with high prevalence and incidence in early age causing great socioeconomic impact, because of both the associated clinical characteristics and treatment [2].

AxSpA involves primarily the entheses of the sacroiliac joints (SIJs) and the spine, which are the most frequently compromised anatomic regions due to this disease. The SIJs are considered to be the most important sites of impairment and magnetic resonance imaging (MRI) is recognized as the most sensitive technique for early diagnosis of inflammatory sacroiliitis due to its great textural contrast resolution, by revealing subchondral bone marrow edema [3].

The Assessment of SpondyloArthritis International Society (ASAS) group recommends T2-weighted MRI sequence sensitive for free water, such as short tau inversion recovery (STIR) or T2 fat saturation (fat-sat), to detect SIJ active inflammation [3]. The MRI characteristics of SIJ related to active inflammation include high-intensity gray levels close to the joint surface, in the subchondral bone, and the depth of that intensity. Figure 1 shows examples of a positive and a negative case for active inflammatory sacroiliitis.

Despite efforts to standardize the evaluation, the qualitative and semiquantitative diagnosis performed by expert radiologists and rheumatologists still remains subject to significant intrapersonal and interpersonal variation [4]. Therefore, this is an important field for potential application of computer-assisted methods using artificial intelligence or machine learning techniques to achieve reliable and early diagnosis.

Machine learning is a branch of artificial intelligence, which allows the extraction of meaningful patterns from examples [5, 6]. The artificial intelligence approach has been widely used in medical image classification tasks, such as melanoma [7], discrimination of smoking status based on deep learning with MRI [8], classification of dermatological ulcers [9], evaluation of breast cancer [10], lung diseases [11, 12], and vertebral compression fractures [13, 14]. Computer-assisted analysis can be based on different approaches, such as statistical methods, instance-based analysis, decision trees, and artificial neural networks (ANNs). However, machine-learning models could have some limitations, for instance, bias to the majority class with imbalanced datasets and overfitting due to high feature-vector dimensionality. Therefore, it is required to evaluate the performance of machine learning techniques for each specific application.

In this context, our proposal was to evaluate the applicability of classical machine learning models and feature selection methods for the classification of active inflammatory sacroiliitis in magnetic resonance images.

Material and methods

This retrospective study was approved by the Institutional Review Board (IRB) at the University Hospital. IRB waived the requirement to obtain informed consent of patients.

Image acquisition and preprocessing

Images from SIJ MRI exams of 56 patients were retrospectively recovered from the Picture Archiving and

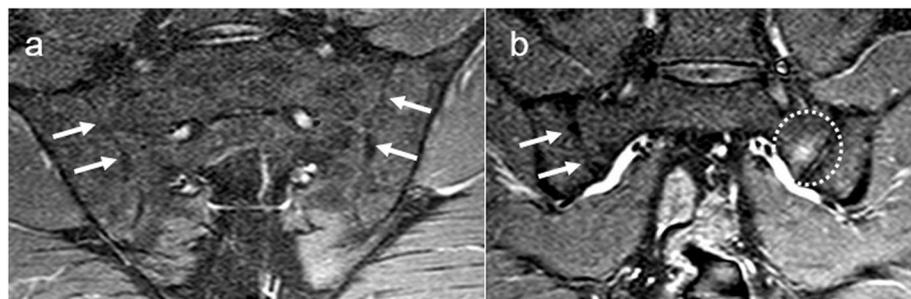


Fig. 1 a. Negative case for active inflammatory sacroiliitis on MRI illustrated with one of its coronal STIR images. There are no hyperintense foci at the subchondral bone adjacent to the articular surfaces (white arrowheads). **b.** A positive example of bone marrow edema related to active sacroiliitis on MRI. The subchondral bone marrow edema is characterized by ill-defined foci of hyperintensity and is shown inside the dotted white circle. White arrowheads indicate the right sacroiliac joint surface

Communication System (PACS) of the University Hospital. Exams were acquired with a 1.5 T scanner (Achieva, Philips Medical Systems), using the spine coil, with the acquisition of coronal STIR sequences. From each MRI exam, a musculoskeletal radiologist selected six images as being the most representative images of the SIJs of the patient, resulting in a total of 336 images.

Patients whose MRIs were included in this study were all initially investigated for suspected inflammatory sacroiliitis. Some of them finally had the diagnosis of spondyloarthritis, and others did not. At the end of 2 years of follow-up, all patients in the positive group (SIJ active inflammation) were diagnosed with spondyloarthritis according to clinical and laboratory criteria. In the negative group (SIJ without active inflammation), half of the patients (13 individuals) did not meet the clinical and laboratorial criteria for spondyloarthritis, and received other diagnosis, such as osteoarthritis, fibromyalgia, gout, or psychiatric disorder. The other half of patients in the negative group, despite having the final diagnosis of spondyloarthritis during follow-up, did not present active inflammation at the time of the MRI examination.

All images were anonymized and manually segmented by the same musculoskeletal radiologist. The segmentation was performed using Adobe Photoshop CC version 14.1 × 64. Two musculoskeletal radiologists classified, in consensus, each MRI exam as positive or negative for active inflammation. One of the radiologists had, at the time of this study, 2 years of experience after a clinical fellowship in musculoskeletal radiology, and the other was a senior radiologist with 18 years of clinical experience. MRI exams were categorized by the radiologists as a positive or a negative test for inflammatory sacroiliitis for each patient according to the ASAS criteria [3]. The MRI criteria used to define positivity of SIJ inflammation correspond to foci of subchondral edema seen at two

different sites or at the same site in at least two consecutive images [3].

The radiologists' classification defined 24 patients as positive and 32 as negative for inflammatory sacroiliitis, and this classification was used as the reference standard to calculate sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC). The dataset was randomly split with ~80% (46 samples, 20 positive and 26 negative) for training and ~20% for external test (10 samples, 4 positive and 6 negative).

Each original image has the spatial resolution of 256×256 pixels and contrast resolution of 256 Gy levels. The SIJ region of interest (ROI) was placed on a black background during the process of manual segmentation. However, this background could cause noise and artifacts in the feature extraction step (described in Section 2.2), such as high frequencies present in the transition between the ROI and the background. To minimize high-frequency noise, a preprocessing method based on the warp perspective transform, including a polynomial transformation [15], was used to expand the ROI and cover all of the background (Fig. 2).

Feature extraction and selection

Statistical analysis was performed based on features extracted from the histograms of the preprocessed ROIs with 256 bins. The features derived included the Mean, Variance, Standard Deviation, Kurtosis, Coefficient of Variation, Skewness, and Maximum Pixel Value.

Texture analysis was based on the features proposed by Haralick et al. [16] using the gray-level cooccurrence matrix, and the features proposed by Tamura et al. [17] extracted from image gray levels. Haralick's features were calculated using a cooccurrence matrix with distance 1 and are listed as follows: Second Angular Momentum, Contrast, Correlation, Variance, Moment of Inverse Difference, Mean Sum, Sum Entropy, Sum of

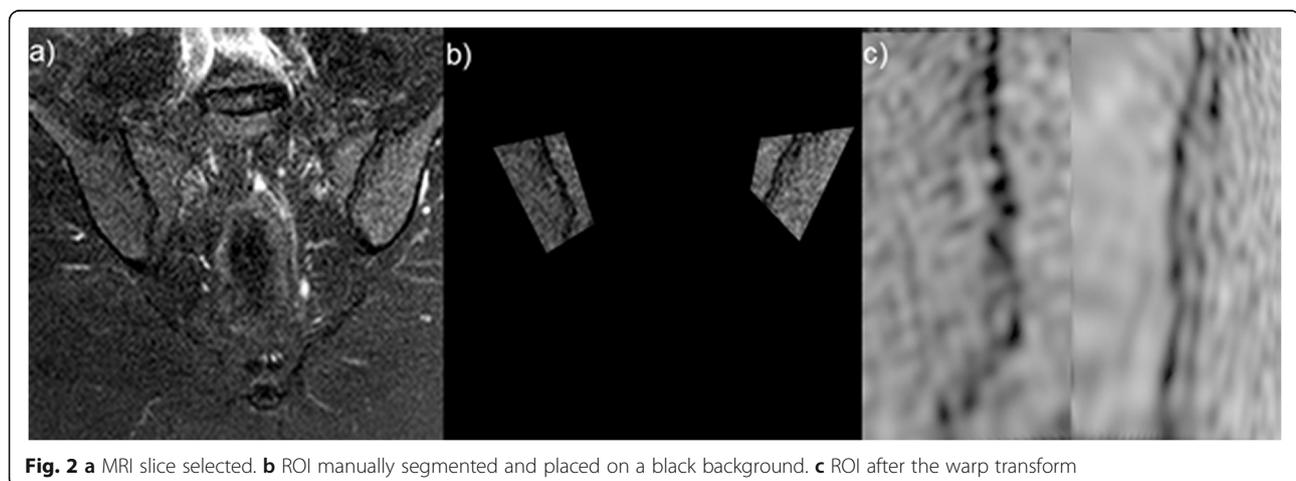


Fig. 2 a MRI slice selected. b ROI manually segmented and placed on a black background. c ROI after the warp transform

Variance, Difference of Variance, Difference of Entropy, two Measures of Information Correlation, and Maximum Correlation Coefficient. Tamura's features were Contrast, Granularity, and Directionality, computed in 16 directions, for a total of 18 features. All of these features were computed using the open-source Java library JFeatureLib [18].

The fast Fourier transform (FFT) was applied to the warped images to obtain the power spectrum using the open-source library ImageJ [19]. The attributes extracted from the two-dimensional rectangular power spectrum were called Fourier features, which include the Mean, Variance, Standard Deviation, Asymmetry, Kurtosis, Coefficient of Variation, and Maximum Pixel Value. The statistics of the power spectrum summarize the frequency intensities, which may be a simple and intuitive way to discriminate instances using frequency features.

The Haar wavelet transform [20] was applied to decompose each image into subimages to obtain the energy in the low-frequency band (LL) and high-frequency bands (HH, HL, LH) in levels 2 and 3. The Haar wavelet is defined as a noncontinuous function and its application to an image results in subimages with vertical, horizontal, and diagonal details from the original image. The energy of each subimage was defined as the sum of all pixel values. The Haar wavelet was implemented using the Fractional Wavelet Module in ImageJ [19].

Gabor filters were applied to each image to obtain the energy in each frequency band, capturing local frequency features [21] in five scales and six orientations. Gabor filters are defined as continuous functions that can detect features in various directions, but have an implicit assumption that all of the images are captured in the same orientation. For each filter output, the mean and standard deviation were calculated, resulting in 60 Gabor features. Gabor filters were implemented using the open-source Java library JFeatureLib [18].

The estimation of fractal dimension was implemented using the box counting method. This approach uses boxes of different sizes and counts the number of occurrences of a specified pattern in the image. Square boxes with width from 6 to 24 pixels were used; boxes with inside pixel values of 50, 100, 150, and 200 were counted; and the mean values of such counts were obtained. The fractal dimension was then estimated as the slope of the line when the logarithm of the mean number of boxes is plotted on the Y axis against the size of the boxes on the X axis. The fractal dimension estimation results in one feature.

The final feature vector for each patient was created using the mean and standard deviation of each feature across the six warped MRI ROIs, because the inflammatory pattern may not be presented in all images of a given exam. This resulted in a 230-dimension feature vector for each patient's MRI exam.

Before classification, all features were normalized to the interval [0,1].

The large dimension of the feature vector defined as above may result in poor performance by the classifiers used in machine learning, a problem known as the curse of dimensionality; therefore, we used two feature selection methods to remove irrelevant or redundant features and reduce the vector dimensionality: ReliefF and Wrapper. The ReliefF method assigns a probability of relevance to each feature based on its individual values between multiple nearest instances [22]. The ReliefF algorithm used in this work was implemented with the Weka machine learning platform [23] using 10 nearest neighbors and a search method based on the Ranker algorithm. The Ranker algorithm sorts the feature list from the highest probability to the lowest.

The Wrapper method uses a learning scheme to select features. The idea behind the Wrapper method is to run the chosen classifier with subsets of the feature vector, evaluate the classifier, and choose the feature set with the highest performance [24]. The classifiers used to select features in this work are the Support Vector Machine (SVM), the Multilayer Perceptron (MLP), and the Instance-Based Algorithm (IBA), which will be explained in Section 2.3. The methods were trained and validated using 10-fold cross-validation and the training dataset (46 samples). A feature was considered to be relevant in the training step if it appeared as relevant in at least two folds. The trained model with the best performance was then tested (external validation) using the test dataset (10 samples). Figure 3 shows a flowchart representation of the experiments carried out.

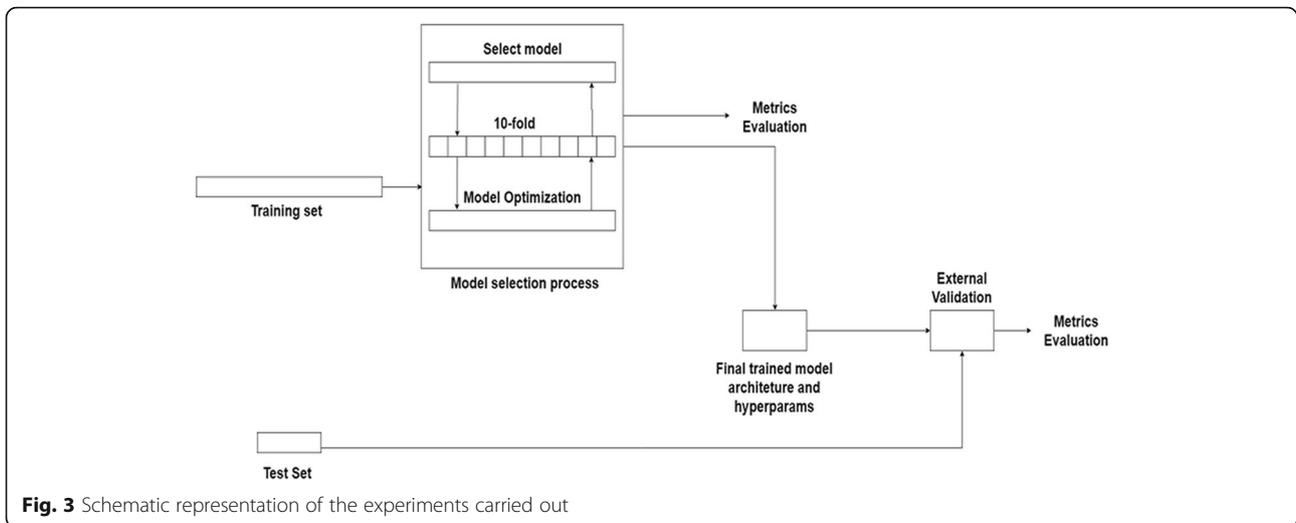
Machine learning

Three machine learning models were used to evaluate the capability of the features to classify SIJ cases.

The SVM is a method that uses hyperplanes to separate the samples provided in an optimal way, such that the margin of separation between the classes (Positive and Negative) will be the maximum possible. The method transforms a multisolution problem to a problem with a unique solution [25]. The equations used to create the hyperplanes were specified to be linear in this work.

IBAs derived from k-nearest neighbors (kNN), referred to as IBk, support robust learning with noisy data, storage reduction during the learning process, and are intuitive [26]. The k values used in this work were 1, 3, and 5.

The MLP is a fully connected ANN which uses backpropagation as the learning scheme. It is a robust model which adjusts synaptic weights according to the error gradient calculated from each training epoch [27]. The MLP model used in this work has the learning rate of 0.3, momentum of 0.2, 500 epochs, and one hidden layer with 231 neurons. The



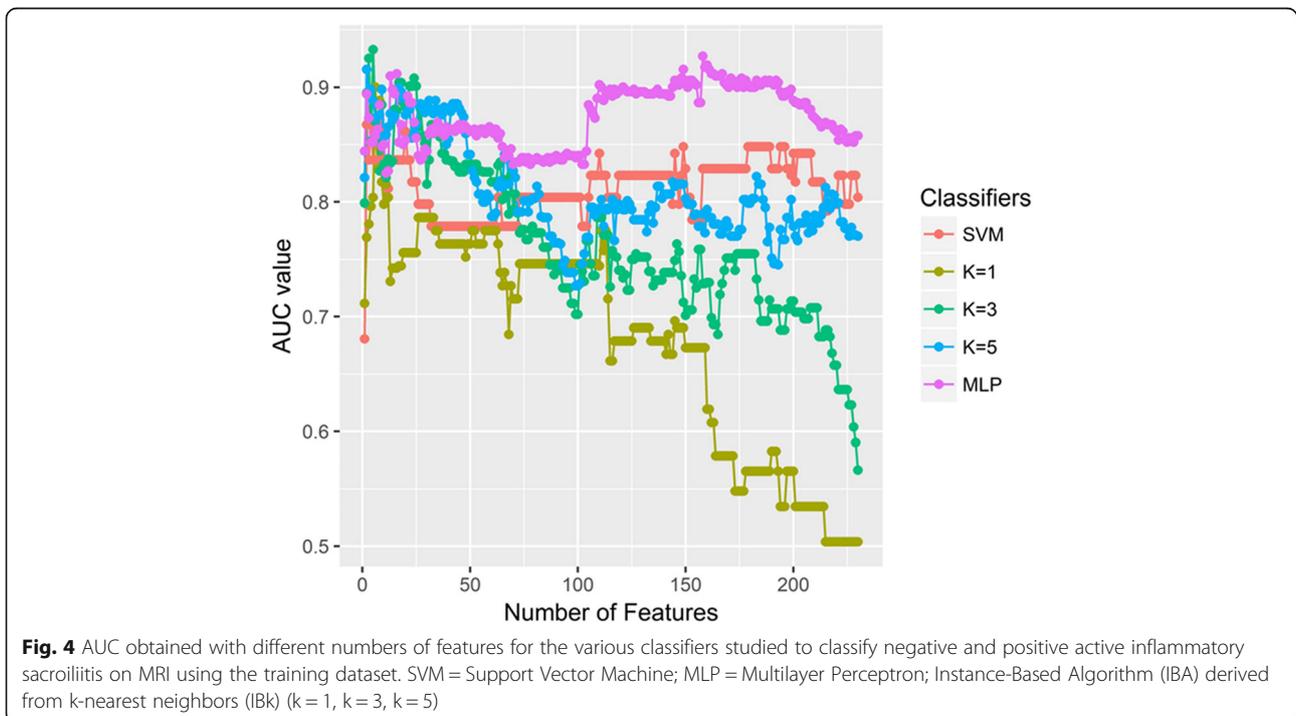
classifiers were implemented using the open-source library Weka [23].

Results

Initial evaluation of the methods was performed using the training dataset and features ranked by ReliefF and Wrapper, which means that we classified the images using N features for each evaluation, where $0 < N < 231$. The results are presented in Figs. 4, 5 and 6 for AUC, sensitivity (true-positive rate), and specificity (true-negative rate), respectively. Table 1

summarizes the best performance of each classifier. Table 2 shows the classification performance using each of the feature vectors selected by the wrapper method for each classifier. Table 3 shows the MLP classifier’s best performance using six features selected by the Wrapper method for the training dataset (10-fold cross-validation, 46 samples) and for the test dataset (external validation, 10 samples).

MLP obtained the best results when all patients categorized as positive for SIJ active inflammation were correctly identified (sensitivity=1). Of the 26 negative



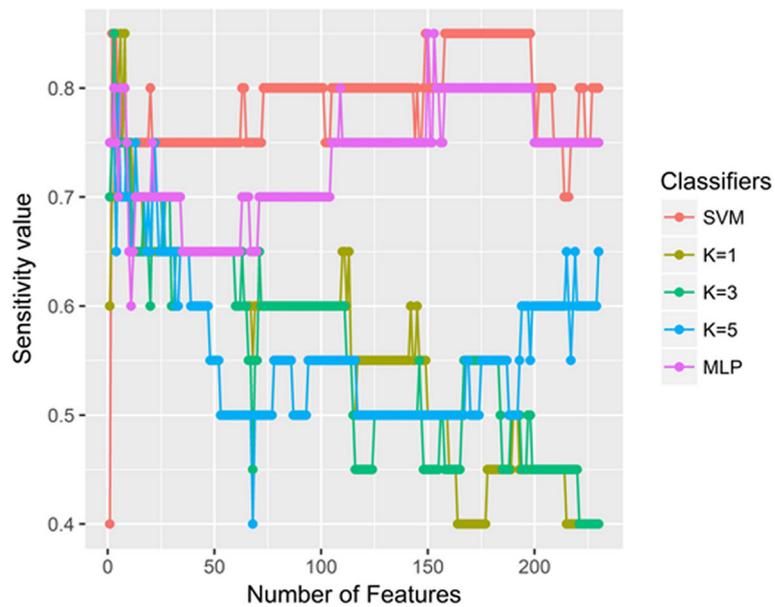


Fig. 5 Sensitivity obtained with different numbers of features for the various classifiers studied to classify negative and positive active inflammatory sacroiliitis on MRI using the training dataset. SVM = Support Vector Machine; MLP = Multilayer Perceptron; Instance-Based Algorithm (IBA) derived from k-nearest neighbors (IBk) (k = 1, k = 3, k = 5)

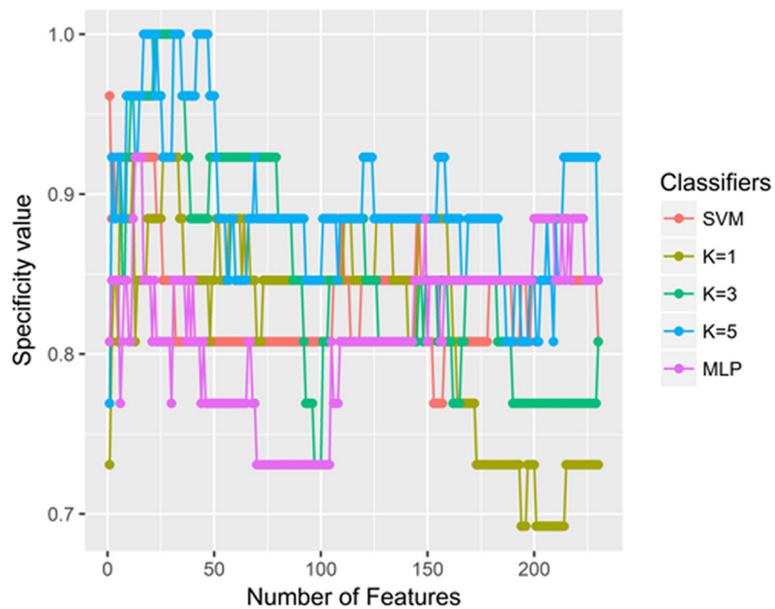


Fig. 6 Specificity obtained with different numbers of features for the various classifiers studied to classify negative and positive active inflammatory sacroiliitis on MRI using the training dataset. SVM = Support Vector Machine; MLP = Multilayer Perceptron; Instance-Based Algorithm (IBA) derived from k-nearest neighbors (IBk) (k = 1, k = 3, k = 5)

Table 1 Best performance for each classifier and number of features used to yield the same result. SVM = Support Vector Machine; MLP = Multilayer Perceptron; Instance-Based Algorithm (IBA) derived from k-nearest neighbors (IBk) (k = 1, k = 3, k = 5); AUC = Area under the ROC curve

Classifier	Metric name	Metric value	Number of features
SVM	AUC	0.867	2
	Sensitivity	0.850	2
	Specificity	0.960	1
IBk with k = 1	AUC	0.900	6
	Sensitivity	0.850	6
	Specificity	0.923	6
IBk with k = 3	AUC	0.932	5
	Sensitivity	0.850	3
	Specificity	1.000	24
IBk with k = 5	AUC	0.915	2
	Sensitivity	0.800	5
	Specificity	1.000	17
MLP	AUC	0.926	158
	Sensitivity	0.85	150
	Specificity	0.923	16

cases, only 2 cases were erroneously classified by the algorithm as positive (specificity = 0.923). Therefore, the final agreement between the radiologists and the algorithm reached 95.6% (Accuracy) in this scenario.

Discussion

Recent literature dedicated to musculoskeletal radiology shows increasing interest in the application of machine learning and other computer techniques, for example, in the analysis of benign and malignant vertebral compression fractures [28], skeletal maturity [29], and differentiation between benign and malignant cartilaginous bone tumors [30]. However, to our best knowledge, there is no previous study dedicated to SpA SIJ inflammation.

Our study examined the use of machine learning models to aid in the classification of MRI of SIJs as positive or negative for active inflammation. The visual diagnosis of sacroiliitis in clinical practice consists of the

detection of changes in the gray levels in the tissues close to the SIJ surfaces by a medical specialist, with high signal intensity of subchondral bone indicating active inflammation. Based on this and related observations, we performed statistical, textural, spectral, and fractal analyses to extract features and characterize SIJs for classification.

In general, classifiers provided their best performance with low-dimension feature vectors obtained using ReliefF or Wrapper methods.

ReliefF provides a classifier-independent list of relevant features. Table 1 shows that kNN with k = 3 reached the highest AUC using only 5 features selected by ReliefF. These five features are the mean of the energy of Haar wavelet for LH on level 2, mean of the maximum value of pixel, standard deviation of the energy of Haar wavelet for HH on level 2, standard deviation for the maximum value of pixel, and standard deviation of the energy of Haar wavelet for LH on level 2.

The high-frequency filters detect abrupt transitions between gray levels. As shown by the results, the maximum value of pixel discriminates between positive and negative instances, indicating that the maximum values are probably causing some high-frequency components in the SIJ ROIs.

For the Wrapper method using 10 folds, kNN with k = 5 reached the highest performance with 9 features. The Wrapper method provides a classifier-based list of relevant features, which are the standard deviation of 6° directionality of Tamura (relevant on 2 folds), standard deviation of 13° directionality of Tamura (relevant on 3 folds), mean of Tamura correlation (relevant on 4 folds), mean of maximum correlation coefficient of Haralick (relevant on 2 folds), mean of the maximum pixel value (relevant on 2 folds), mean of skewness of the Fourier power spectrum (relevant on 3 folds), mean of Haar wavelet from LL on level 2 (relevant on 2 folds), mean of Haar wavelet of LH on level 2 (relevant on 8 folds), and mean of fractal dimension (relevant on 2 folds).

Again, the maximum pixel value and Haar wavelet from LH on level 2 were selected as relevant, implying that these features are, in fact, discriminative. High-frequency components are caused by large changes in

Table 2 - Performance of each classifier using the features selected by the wrapper method. SVM = Support Vector Machine; MLP = Multilayer Perceptron; Instance-Based Algorithm (IBA) derived from k-nearest neighbors (IBk) (k = 1, k = 3, k = 5); AUC = Area under the ROC curve

Classifier	AUC	Sensitivity	Specificity	Accuracy (%)	Number of Features
SVM	0.842	0.800	0.885	87.8	15
IBk with k = 1	0.798	0.800	0.769	78.2	13
IBk with k = 3	0.867	0.750	0.885	82.6	14
IBk with k = 5	0.969	0.750	0.962	86.9	9
MLP	0.965	1.000	0.923	95.6	6

Table 3 - Multilayer perceptron (MLP) classifier selected model performance using 10-fold cross-validation on training samples (46 samples) and external validation on test set (10 samples)

Model	10-Fold			Test set		
	Sensitivity	Specificity	Accuracy (%)	Sensitivity	Specificity	Accuracy (%)
MLP	1	0.923	95.6%	1	0.667	80%

gray levels across small distances in the image. However, not all frequency features were selected, which is probably due to correlation between those features, a characteristic that is detected by the Wrapper method.

The classifier that reached the highest AUC, kNN, has a performance problem in prediction because it always needs to calculate the distance between the predicted instance and all other instances, which is not scalable. If scalability is important, the MLP may be the better choice of classifier using the Wrapper method. The features selected by the MLP are the mean of the 1° directionality of Tamura (relevant on 2 folds), standard deviation of the 13° directionality of Tamura (relevant on 2 folds), mean of sum variance from the gray levels (relevant on 4 folds), mean of maximum pixel value (relevant on 2 folds), mean of second Gabor directionality (relevant on 5 folds), and mean of Haar wavelet from LH on level 2 (relevant on 10 folds).

An important observation is that, always, the maximum pixel value and Haar wavelet from LH on level 2 were selected as relevant by both feature selection methods, asserting that these features are important to discriminate instances. The maximum pixel value is intuitive to be important because inflammation manifests as high-intensity of gray level around the SIJ. Gabor directionality measures are probably selected due to the directionality change caused by the depth of inflammation in SIJ.

We have used the STIR sequence in the coronal plane to apply the machine learning methods, but in clinical practice, radiologists may have access to other fluid-sensitive fat-saturated MRI sequences with images acquired also in the axial and sagittal planes. We chose to use the STIR sequence because this is one of the recommended sequences by the ASAS guidelines [3]. Recently, two different studies have shown that other fluid-sensitive fat-saturated MRI techniques may be equally sensitive and accurate in the diagnosis of inflammatory sacroiliitis [31, 32]. Therefore, it could be interesting to investigate in the future if different fluid-sensitive fat-saturated MRI techniques could provide and support similar results using the machine learning approach. We also encourage future studies exploring the potential of radiomics [33, 34] in the evaluation of inflammatory sacroiliitis, with the potential impact of deriving new diagnostic and prognostic information.

We did not investigate the potential of artificial intelligence techniques to identify postinflammatory structural damage on the SIJ surface and subchondral bone, because the aim was to classify active inflammation. However, the identification of such abnormalities may be important for the diagnosis of SpA and future studies should use T1-weighted sequences for this assessment, since these sequences provide a greater conspicuity of such findings.

As expected, in the external validation test the accuracy fell down from 95.6 to 80.0 and specificity from 0.923 to 0.667 (Table 3). We believe that our results are still encouraging, and we suggest new studies to improve AI techniques to investigate inflammatory sacroiliitis and SpA.

Some limitations of this study need mentioning. First, the study was retrospective. In addition, the number of patients was relatively small, which usually precludes the use of deep learning methods [35]. We used the segmentation performed by only one musculoskeletal radiologist as the ground truth, but even experienced specialists may show interpersonal variability, and the inclusion of more radiologists would be desirable to validate future artificial intelligence algorithms. Besides, to use the methods described in our study, it is necessary that a musculoskeletal radiologist or an experienced rheumatologist choose the most representative images of the synovial SIJ region on the coronal plane. The development of a semiautomatic or automatic segmentation tool would be desirable to obviate this workload. Finally, although the database was divided into training and testing sets, which made it possible to make an independent evaluation of the generalization of the validated classifier model obtained during the training phase (10-fold cross-validation), our study enrolled cases from only one institution. Future validation with cases from another institution is required before one can generalize our results for potential clinical application.

Conclusion

Our results show the potential of machine learning methods to identify SIJ subchondral bone marrow edema in axSpA patients and are promising to aid in the detection of active inflammatory sacroiliitis on MRI STIR sequences. Multilayer Perceptron (MLP) achieved the best results.

Acknowledgments

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

Authors' contributions

MCF and PMAM conceived and designed the experiments; MCF, APMT, VFD, RLA analyzed the data, MCF and MHN-B drafted the manuscript; MCF, JRFJ performed experiments and acquired data; MCF, VFD and RLA collected samples; MCF, PMAM, RMR, PL-J, MHN-B interpreted data and critically revised the manuscript. All authors read and approved the final version of the manuscript.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001; by the São Paulo Research Foundation (FAPESP) 2016/17078-0 and 2018/07765-6; by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) 452257/2018-2 and 305124/2018-8. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The Ethics Committee of our University Hospital approved this research project, process number 1.600.012 and CAAE: 56887916.0.0000.5440.

Consent for publication

All authors read and approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

Author details

¹São Carlos School of Engineering, University of São Paulo, São Carlos, SP, Brazil. ²Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil. ³MAInLab Medical Artificial Intelligence Laboratory, Ribeirão Preto Medical School, Ribeirão Preto, Brazil. ⁴Ribeirão Preto Medical School Musculoskeletal Imaging Research Laboratory, Ribeirão Preto, Brazil. ⁵Radiology Division / CCIFM, Ribeirão Preto Medical School, Av. Bandeirantes, 3900, Ribeirão Preto, SP CEP 14048-900, Brazil. ⁶Electrical and Computer Engineering Schulich School of Engineering University of Calgary, Calgary, Alberta, Canada.

Received: 10 September 2019 Accepted: 16 April 2020

Published online: 07 May 2020

References

- Garg N, van der Bosh F, Deodhar A. The concept of Spondyloarthritis: where are we now? *Best Pract Res Clin Rheumatol*. 2014;28:663–72.
- Boonen A. Socioeconomic consequences of ankylosing spondylitis. *Clin Exp Rheumatol*. 2002;20:S23–6.
- Lambert RGW, Bakker PAC, van der Heijde D, Weber U, Rudwaleit M, et al. Defining active sacroiliitis on MRI for classification of axial spondyloarthritis: update by the ASAS MRI working group. *Ann Rheum Dis*. 2016;75:1958–63.
- Maksymowich WP, Inman RD, Salonen D, Dhillon SS, Williams M, Stone M, Conner-Spady B, Palsat J, Lambert RGW. Spondyloarthritis research consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in Ankylosing spondylitis. *Arthritis Rheumatism*. 2005;53:703–9.
- Erickson BF, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *RadioGraphics*. 2017;37:505–15.
- Chartrand G, Cheng PM, Vorontsov E, Drozdal M, Turcotte S, Pal CJ, Kadoury S, Tang A. Deep learning: a primer for radiologists. *Radiographics*. 2017;37:2113–31.
- Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Jafari MH, Ward K, Najarian K. Melanoma detection by analysis of clinical images using convolutional neural network. *Conf Proc IEEE Eng Med Biol Soc*. 2016;2016:1373–6.
- Wang S, Zhang R, Deng Y, Chen K, Xiao D, Peng P, Jiang T. Discrimination of smoking status by MRI based on deep learning method. *Quant Imaging Med Surg*. 2018;8:1113–20.
- Pereira SM, Frade MAC, Rangayyan RM, Azevedo-Marques PM. Classification of color images of dermatological ulcers. *IEEE Journal of Biomedical and Health Informatics*. 2013;17:136–42.
- Azevedo-Marques PM, Rosa NA, Traina AJM, Traina Junior C, Kinoshita SK, Rangayyan RM. Reducing the semantic gap in content-based image retrieval in mammography with relevance feedback and inclusion of expert knowledge. *Int J Comput Assist Radiol Surg*. 2008;3:123–30.
- Ferreira Junior JR, Koenigkam-Santos M, Cipriano FER, Fabro AT, Azevedo-Marques PM. Radiomics-based features for pattern recognition of lung cancer histopathology and metastases. *Comput Methods Prog Biomed*. 2018;159:23–30.
- Allende-Cid H, Rangayyan RM, Azevedo-Marques PM, Almeida E, Frery A, Cardoso I, Ramos H. Analysis of machine learning algorithms for diagnosis of diffuse lung diseases. *Methods Inf Med*. 2018;57:272–9.
- Azevedo-Marques PM, Spagnoli HF, Frighetto-Pereira L, Reis RM, Metzner GA, Rangayyan RM, Nogueira-Barbosa MH. Classification of vertebral compression fractures in magnetic resonance images using spectral and fractal analysis. *Conf Proc IEEE Eng Med Biol Soc*. 2015;2015:723–6.
- Casti P, Mencattini A, Nogueira-Barbosa MH, Frighetto-Pereira L, Azevedo-Marques PM, Martinelli E, Di Natale C. Cooperative strategy for a dynamic ensemble of classification models in clinical applications: the case of MRI vertebral compression fractures. *Int J Comput Assist Radiol Surg*. 2017;12:1971–83.
- Faleiros MC, Zavala EJR, Ferreira-Junior JR, Dalto VF, Assad RL, Louzada Junior P, Nogueira-Barbosa MH, Azevedo-Marques PM. Computer-aided classification of inflammatory sacroiliitis in magnetic resonance imaging. *Int J Comput Assist Radiol Surg*. 2017 Jun;12(Suppl 1):154–5.
- Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Transactions on Systems man and cybernetics*. 1973;SMC-3:610–523.
- Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man, & Cybernetics*. 1978;8:460–73.
- Keuschning M & Penz C. JFeatureLib open source project. 2008. Available in <http://github.com/locked-fg/JFeatureLib>. Accessed 1 May 2020.
- Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9:671–5.
- Haar A. Zur Theorie der orthogonalen Funktionensysteme. *Math Ann*. 1910;69:331–71.
- Zhang D, Wong A, Indrawan M, Lu G. Content-based image retrieval using Gabor texture features. University of Sydney: IEEE Pacific-Rim Conference on Multimedia; 2000.
- Kononenko I. Estimating attributes: analysis and extensions of relief. *European Conference of Machine Learning*; 1994.
- Frank E, Hall MA, Witten I. In: Kaufmann M, editor. *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*. 4th ed; 2016.
- Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell*. 1997;97:273–324.
- Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intelligent Systems and their Appl*. 1998;13:18–28.
- Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. *Mach Learn*. 1991;6:37–66.
- Haykin S. *Neural networks - a comprehensive foundation* second edition, Pearson education; 1999.
- Frighetto-Pereira L, Rangayyan RM, Metzner GA, Azevedo-Marques PM, Nogueira-Barbosa MH. Shape, texture and statistical features for classification of benign and malignant vertebral compression fractures in magnetic resonance images. *Comput Biol Med*. 2016;73:147–56.
- Larson DB, Chen MC, Lungren MP, Halabi SS, Stence NV, Langlotz CP. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. *Radiology*. 2018;287:313–22.
- Lisson CS, Lissou CG, Flodorf K, Mayer-Steinacker R, Schultheiss M, von Baer A, Barth TFE, Beer AJ, Baumhauer M, Meier R, Beer M, Schmidt SA. Diagnostic value of MRI-based 3D texture analysis for tissue characterisation and discrimination of low-grade chondrosarcoma from enchondroma: a pilot study. *Eur Radiol*. 2018;28:468–77.

31. Dalto VF, Assad RL, Crema MD, Louzada-Junior P, Nogueira-Barbosa MH. MRI assessment of bone marrow oedema in the sacroiliac joints of patients with spondyloarthritis: is the SPAIR T2w technique comparable to STIR? *Eur Radiol.* 2017;27:3669–76.
32. Sung S, Kim HS, Kwon JW. MRI assessment of sacroiliitis for the diagnosis of axial spondyloarthritis: comparison of fat-saturated T2, STIR and contrast-enhanced sequences. *Br J Radiol.* 2017;90(1078):20170090.
33. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* 2016;278:563–77.
34. Hou Z, Yang Y, Li S, Yan J, Ren W, Liu J, Wang K, Liu B, Wan S. Radiomic analysis using contrast-enhanced CT: predict treatment response to pulsed low dose rate radiotherapy in gastric carcinoma with abdominal cavity metastasis. *Quant Imaging Med Surg.* 2018;8:410–20.
35. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;28:436–44.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

