We implemented automated hyper-parameter tuning software into the dedicated login node. The training jobs using the hyper-parameters generated by the software were repeatedly executed at the compute nodes. We utilized Bayesian optimization (BO) [1] as an automated hyper-parameter tuning algorithm. BO is a framework for an optimization of the black-box functions whose derivatives and convexity properties are unknown. In training of deep learning using our framework, the hyper-parameters were chosen by BO so as to maximize the value of evaluation criteria in validation.

Results

We trained CAD using deep learning in the constructed environment. We targeted a computerized detection of cerebral aneurysm in magnetic resonance angiography (MRA) images based on a deep convolutional neural network (CNN) [2]. We utilized 3D time-of-flight unenhanced MRA data sets of 350 cases accumulated from three 3-Tesla MR scanners (two Signa HDxt and one Discovery MR750, GE Healthcare, Waukesha, WI, USA). The acquisition parameters were as follows: echo time, 2.7–3.3 ms, repetition time, 22 or 25 ms; flip angle, 15 degree; field of view, 240 mm; slice thickness/interval, 0.6/1.2 mm; matrix size, 512 × 512. Each case includes at least one aneurysm of 2 mm or more in diameter, which was determined by consensual reading by two experienced radiologists, and areas of aneurysm were defined by pixel-by-pixel painting. We divided the 350 cases into two subsets; 300 cases of training set, and 50 cases of validation set.

In our CAD, we used a CNN classifier that predicts whether each voxel was inside or outside aneurysms by inputting maximum intensity projection images generated from a volume of interest around the voxel. Our network consisted of two convolutional layers, two max-pooling layers, and two fully-connected layers. The output layer had a single unit, and the logistic function was applied to the output to convert it into the probability of being positive (which ranges from 0 to 1). We employed a rectified linear unit (ReLU) function as the activation function for all layers except the output layer. Batch normalization was performed before each ReLU function. We utilized the Adam method to optimize the network weights. The tuned hyper-parameters were: the filter size and the number of filters of each convolution layer, the number of units of the fully-connected layer, the batch size, and three parameters (α, β_1, β_2) of the Adam method. We utilized the area under the curve (AUC) value of the free-response receiver operating characteristic (FROC) curve, with the upper limit of three false positives per case, as an evaluation criterion. In this study, the 40 trials of hyper-parameter tuning with BO were repeated five times.

Figure 2 shows the change in the maximum AUC value with error bar. The value of each trial indicates the maximum value in all past trials. As the number of trials increases, the maximum AUC value has been updated, indicating that the appropriate hyper-parameters were found by BO.

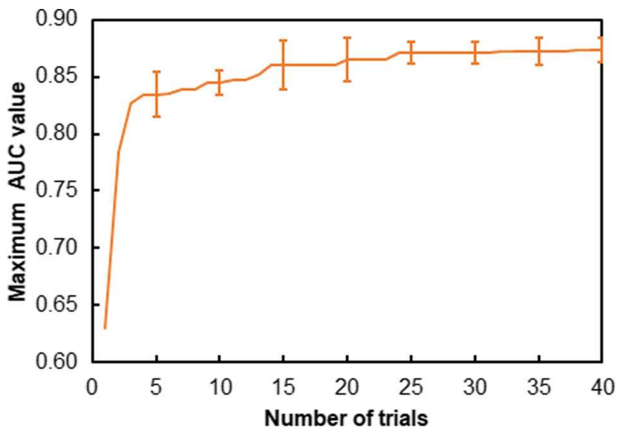

Fig. 2 Change in AUC of validation data where each value is maximum value in past trials

Conclusion

We have built a training environment for deep learning on supercomputer system. The constructed environment enabled to train deep learning model with hyper-parameter tuning. We are planning to validate our environment using other type of CAD, and to implement asynchronous parallel BO algorithm [3] for further efficiency in training.

References

[1] Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. Advances in neural information processing systems: 2951–2959.

[2] Nakao T, Hanaoka S, Nomura Y, Sato I, Nemoto M, Miki S, Maeda E, Yoshikawa T, Hayashi N, Abe O (2017) Deep neural network-based computer-assisted detection of cerebral aneurysms in MR angiography. J Magn Reson Imaging 2017 Aug 24 (Epub ahead of print)

[3] Kandasamy K, Krishnamurthy A, Schneider J, Poczos B (2017) Asynchronous parallel Bayesian optimisation via Thompson sampling. arXiv: 1705.09236

Features selection analysis to quantify sacroiliitis in magnetic resonance imaging

M. Calil Faleiros[1], J. Raniery Ferreira Junior[1],
A. Priscilla Magalhães Tenório[1], V. Faeda Dalto[1], R. Luppino Assad[1],
M. Henrique Nogueira-Barbosa[1],
P. Mazzoncini de Azevedo-Marques[1]
[1]University of São Paulo, Ribeirão Preto, Brazil

Keywords Feature selection · Machine learning · Sacroiliitis · Imagnetic resonance imaging

Purpose

Spondyloarthritis (SpA) is a group of diseases with common clinical and radiological manifestations. SpA comprises mainly the appendicular skeleton, spine, and the sacroiliac joint. SpA affects mostly young adults and may present the first symptoms at 16 years old, potentially impacting morbidity and socioeconomics [1]. The main technique for early diagnosis of sacroiliitis is magnetic resonance imaging (MRI). Computer-aided classification of sacroiliac joints has showed very promising results, potentially aiding the diagnosis of sacroiliitis using quantitative image features and a classical feature selection method of machine learning [2]. However it still presents results with a high dimensional feature vector, which may contribute to a complex model without relevantly increasing the performance of the classification. Therefore, the aim of this work is to combine different methods of feature selection in order to reach a set of features with low dimensionality and high performance with machine learning classifiers.

Methods

Our institutional research board approved this retrospective study with a waiver of patients' informed consent. MRI exam of 51 patients were used in this study after anonymization. Each MRI exam comprised 6 images, each image was manually segmented and put on a black background. A musculoskeletal radiologist classified each image according to Spondyloarthritis Research Consortium of Canada (SPARCC) score. This classification was the reference standard to evaluate AUC, sensitivity and specificity. The previously classification defined 22 patients positive for sacroiliitis and 29 negative. Images were pre-processed by the warp perspective transform to remove the black background, which causes noise to some features [1]. The features extracted from each image were gray-level statistics, textural based on cooccurrence matrix, textural based on histogram, spectral based on frequency domain, spectral based on wavelets and fractal. Each exam was characterized by the mean and standard deviation of each feature for the 6 images, totalizing 230 features.

Three feature selection methods were combined to filter the final vector, initially composed of 230 features. Mann–Whitney *U* test is a statistical method that is related to the area under the receiver operating characteristics (ROC) curve [3], giving a *p* value for each feature according to their statistical significance. ReliefF is a method that assigns a probability of relevance to each feature based on their individual value between multiple nearest instances [4]. Finally, Wrapper is a method that uses classifiers and an incremental learning scheme to select features [5]. The classifiers used to select features with the Wrapper and to classify the images were naive bayes (NB), multilayer perceptron (MLP), decision tree j48 (J48), random forest (RF), and support vector machine (SVM), resulting in a set of features selected by all classifiers.

Due to the class imbalance problem, the dataset samples were balanced using the synthetic minority over-sampling technique (SMOTE) method. Each classifier were evaluated by area under the ROC (AUC), sensitivity, and specificity, using the 10-fold cross-validation method.

Results

The Mann–Whitney *U* test selected 5 features statistically significant ($p < 0.001$), the ReliefF method selected 6 features with probability threshold of 0.05, and the Wrapper method selected 8 features which is common to the 5 classifiers used.

Using a simple intersection of those three feature sets, we found 4 features that are common to the three methods. These features are 3 energies of high-frequency from Haar wavelet and one gray-level statistic, skewness.

Table 1 presents the results of the classification performed by the NB, J48, MLP, RF and SVM classifiers using those 4 features.

Table 1 Results of each metric for each classifier

	AUC	Sensitivity	Specificity
MLP	0.807	0.793	0.793
NB	0.873	0.828	0.759
J48	0.616	0.655	0.586
RF	0.817	0.793	0.759
SVM	0.828	0.828	0.828

The NB method obtained the highest AUC (0.873), but is less sensitive to predict negative cases than SVM (specificity of 0.759 for NB and 0.828 for SVM). SVM obtained the same value of sensitivity as the NB method (0.828). However, the AUC for SVM (0.828) was close to the AUC of NB (0.873), suggesting the results are statistically the same or very similar.

Conclusion

This work used three methods to select features among a set of gray-level statistical, textural, spectral and fractal. Machine learning analysis was performed to evaluate this feature set efficiency to classify MRI sacroiliitis.

The classification showed that the low dimensional feature vector may be a good approach to classify inflammatory sacroiliitis. Features of gray-level statistics (skewness) and Haar wavelets (3 s level high-frequency energies) have showed efficiency to perform sacroiliitis classification. We propose for future work to use deep learning methods to perform the classification without any previously feature extraction.

References

[1] Costa IP, Bortoluzzo AB, Gonçalves CR, Silva JAB, Ximenes AC, Bértolo MB, Ribeiro SLE, Keiserman M, Menin R, Skare TL, Carneiro S, Azevedo VF, Vieira WP, Albuquerque EN, Bianchi WA, Bonfiglioli R, Campanholo C, Carvalho HMS, Pinto Duarte ALB, Kohem CL, Leite NH, Lima SAL, Meirelles ES, Pereira IA, Pinheiro MM, Polito E, Resende GG, Rocha FAC, Santiago MB, Sauma MFLC, Valim V, Sampaio-Barros PD (2015) Avaliação do desempenho do BASDAI (Bath Ankylosing Spondylitis Disease Activity Index) numa coorte brasileira de 1.492 pacientes com espondiloartrites: dados do Registro Brasileiro de Espondiloartrites (RBE). Revista Brasileira de Reumatologia 55(1): 48–54.

[2] Faleiros MC, Ferreira Junior JR, Jens EJ, Dalto VF, Nogueira-Barbosa MH, Azevedo-Marques PM (2017) Reconhecimento computadorizado de padrões inflamatórios de sacroiliíte em imagens de ressonância magnética. XXXVII Congresso da Sociedade Brasileira de Computação. 17° WIM—Workshop de Informática Médica: 1845–1848.

[3] Mason SJ, Graham NE (2002) Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. Quarterly Journal of the Royal Meteorological Society 128: 2145–2166.

[4] Kononenko I (1994) Estimating Attributes: Analysis and Extensions of Relief. European Conference of Machine Learning: 171–182.

[5] Kohavi R, John GH (1997) Wrappers for feature subset selection, Artificial Intelligence 97(1–2): 273–324.

Radiomics association of quantitative CT features with lung cancer patterns

J. R. Ferreira Junior[1], M. Koenigkam-Santos[2], A. Magalhães Tenório[2], M. Calil Faleiros[1], F. Garcia Cipriano[2], A. Todorovic Fabro[2], J. Näppi[3], H. Yoshida[3], P. M. Azevedo-Marques[2]

[1]Universidade de São Paulo, Programa de Pós-Graduação Interunidades em Bioengenharia, São Carlos, Brazil

[2]University of São Paulo, Ribeirão Preto Medical School, Ribeirão Preto, Brazil

[3]Harvard Medical School, Massachusetts General Hospital, Boston, United States

Keywords Lung cancer · Radiomics · Quantitative image analysis · Medical image computing