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

Purpose

The prognosis of lung carcinoma, the deadliest of all cancers, varies markedly according to tumor staging at diagnosis [1]. One of the most important prognostic factors used to determine therapy is the tumor-node-metastasis (TNM) staging system. Another important prognostic factor is the pathological subtype of the tumor [2]. The two most common lung cancer subtypes are adenocarcinoma (ADC) and squamous cell carcinoma (SCC). Computed tomography (CT) features have also been used as predictive factors, influencing prognosis and response to therapy [1]. Recently, radiomics has emerged as a quantitative imaging approach to improve diagnosis and assessment of prognosis, thereby providing decision support for precision medicine. Radiomics involves computer-based extraction of image features to combine them with other patient data for establishing associations with clinical outcomes [3]. In this work, we aim to extract several quantitative 2D and 3D CT image features, combine them with patient and tumor clinical data, and associate those with tumor pathology and presence of nodal and distant metastases, to potentially aid the pathological diagnosis and treatment decision of lung cancer based on metastases status.

Methods

Our institutional review board approved this retrospective study with a waiver of patients' informed consent. Our lung cancer image cohort has 85 malignant lung tumors with pathology confirmed by biopsy or surgical resection. The presence of metastases was assessed according to the clinical TNM staging system. Patients were imaged using thin-slice contrast-enhanced CT. The lung tumors were semi-automatically segmented from the CT images using the volumetric region growing algorithm GrowCut [4].

Tumors were characterized by 2465 quantitative image features extracted from the segmented CT regions/volumes of interest, including features of gray-level intensity, histogram, cooccurrence matrix, run-length matrix, neighborhood intensity-difference matrix, Tamura texture, Laplacian of Gaussian filters, Gabor filters, Fourier transform, Haar wavelet, fractal dimension, and shape.

The image features were first assessed univariately by the Mann–Whitney *U* test across the different tumor patterns for the pathological subtype (ADC or SCC), presence of nodal metastasis (NM+ or NM−), and presence of distant metastasis (DM+ or DM−). The image features were then normalized and combined with the clinical data of patient's age, gender, smoker status, presence of other primary tumors on patient's body, and the lung lobe location, position, and diameter of the tumor. All combined patient information (quantitative image features and clinical data) were assessed multivariately by a machine-learning algorithm based on the method ReliefF for feature selection and a radial basis function network (RBFN) for image classification [4]. The machine-learning algorithm finds the highest classification performance by determining an optimal combination of x selected features and y neurons for the RBFN single hidden layer, where x was varied in the interval [1, 100] and y was varied in the interval [1, number of samples of the majority class]. The machine-learning algorithm was evaluated using the leave-one-out cross-validation method. Due to class imbalance, the majority class for each pattern was randomly undersampled. Association was assessed by use of the area under the receiver operating characteristic curve (AUC).

Results

Table 1 presents the associative performance of quantitative image features and clinical data for different lung cancer patterns, using the statistical univariate and machine-learning multivariate analyses.

Table 1 Highest AUC values obtained by the univariate and multivariate analyses

	Univariate analysis	Multivariate analysis
Pathology (ADC vs. SCC)	0.66 (COM3D Max Probability)	0.87 (32 features, 14 neurons)
Nodal Metastasis (NM+ vs. NM−)	0.76 (Wavelet HL3)	0.78 (2 features, 2 neurons)
Distant Metastasis (DM+ vs. DM−)	0.70 (Wavelet LH2)	0.92 (13 features, 15 neurons)

Feature or number of features and the number of hidden layer neurons that yielded the highest performance are indicated in parentheses

The Max-Probability feature of the 3D gray-level cooccurrence matrix yielded the highest associative performance for pathologic patterns (p value < 0.05). The level-3 high-frequency HL feature of Haar wavelet yielded the highest associative performance for lymph nodal metastatic patterns (p value < 0.001), and the level-2 low-frequency LH feature of Haar wavelet yielded the highest associative performance for distant metastatic patterns (p value < 0.01). However, the machine-learning algorithm outperformed the univariate statistical analysis in discriminating the pathology and distant metastasis (p value < 0.05). Combining the most relevant features (according to the ReliefF feature selection method) with clinical data and the RBFN increased associative performance with an AUC difference by up to 0.22 units (distant metastasis).

Conclusion

In this work, we associated several quantitative first-order, second-order, and higher-order 2D and 3D CT image features with pathological, nodal and distant metastatic patterns of lung cancer. Statistical and machine learning analyses were performed to assess the associative performance of the features across three different types of malignant lung tumor patterns.

The Max-Probability feature of the 3D gray-level cooccurrence matrix presented a high association with pathological subtypes of lung cancer. The HL3 and LH2 features of Haar wavelet presented a high association with nodal and distant metastases of lung cancer, respectively. However, highest associative performances for pathological and metastatic tumor patterns were obtained by combining several different quantitative CT image features with prognostic clinical data and using the machine-learning algorithm. We propose for future directions to associate quantitative image features with additional types of lung cancer patterns and clinical outcomes, such as tumor recurrence, genomic mutations, and patient survival.

Acknowledgements

We thank CAPES, PDSE (Grant #88881.134004/2016-01), CNPq, FAPESP (Grant #2016/17078-0), and FAEPA-HCFMRP-USP for the financial support.

References

[1] Koenigkam-Santos M, Muley T, Warth A, de Paula WD, Lederlin M, Schnabel PA, Schlemmer HP, Kauczor HU, Heussel CP, Puderbach M (2014) Morphological computed tomography features of surgically resectable pulmonary squamous cell carcinomas: Impact on prognosis and comparison with adenocarcinomas. European Journal of Radiology 83(7): 1275–1281.

[2] Cooper WA, O'toole S, Boyer M, Horvath L, Mahar A. (2011) What's new in non-small cell lung cancer for pathologists: the importance of accurate subtyping, EGFR mutations and ALK rearrangements. Pathology 43(2): 103–115.

[3] Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278(2): 563–77.

[4] Ferreira Junior J, Garcia Cipriano FE, Todorovic Fabro, A, Koenigkam-Santos M, de Azevedo-Marques PM (2018) Radiomics-based recognition of metastatic and histopathological patterns of lung cancer. Lecture Notes in Computational Vision and Biomechanics 27: 613–623.

Deep-learning combined with radiomic hyper-curvature features for survival prediction of patients with interstitial lung disease

R. Nasirudin[1], M. Matsuhiro[1,2], C. Watari[1], J. Nappi[1], T. Hironaka[1], Y. Kawata[2], N. Niki[2], H. Yoshida[1]
[1]Massachusetts General Hospital/Harvard Medical School, Radiology, Boston, United States
[2]Tokushima University, Tokushima, Japan

Keywords Deep learning · Radiomics · Computer-aided diagnosis · Interstitial lung diseases

Purpose

Rheumatoid arthritis (RA) is the most common connective tissue disease that develops inflammatory synovitis, which affects approximately 1% of people in the United States [1]. Interstitial lung disease (ILD) is one of the major extra-articular manifestations, where 5–10% of patients with RA suffer from clinically significant ILD. Survival after the diagnosis of RA-associated ILD (RA-ILD) has been reported as a median of 3–8 years. Currently, however, there is no established imaging predictor for the survival of patients with RA-ILD.

The purpose of this study was to evaluate the effect of deep-learning-derived radiomic features obtained from CT images, called deep radiomic features (DRFs), and their combination with CT radiomic features called hyper-curvature features, on the prediction of the overall survival of patients with RA-ILD in comparison with an established clinical prognostic biomarker known as the gender, age, and physiology (GAP) index [2].

Methods

We retrospectively identified 70 RA-ILD patients who underwent thin-section lung CT and pulmonary function tests, from the medical records of our institution. An experienced observer (an internist with 15 years of experience in pulmonary disease diagnosis and treatment) extracted approximately 4600 regions of interest (ROIs) from the CT images of all patients, and labeled them as having one of the following five lung texture patterns: normal, consolidated, ground-glass opacity, reticulation, or honeycombing. Then, an image patch of size 24 × 24 pixels was extracted at each point within these ROIs.

The image patches were subjected to a modified version of a publicly available DCNN (Cifar-10 CNN), which consisted of three layers of convolutional and max-pooling layers, for extracting hierarchical features with increasing levels of abstraction, followed by two fully connected layers for classification of the input images. The top-most layer of the DCNN outputs the probabilities at which the input voxel belongs to each of the five lung texture patterns. A 512-dimensional DRF vector for each patch was identified as an output of the last convolutional layer of the DCNN, and a 5-fold cross-validation method was employed for an unbiased estimate of the DRF vectors. The mean, standard deviation, and maximum of each element of the DRF vectors were computed on the lungs of a patient to derive a 1536-dimensional DRF vector for characterizing the patient.

We also computed CT radiomic features called hyper-curvature features [4] that included 4D principal curvatures, curvedness, blight-sheet, blight-cylinder, blight-blobs, dark-cylinder, dark-line, dark-blob, and the scale thereof on the lungs. Mean of each of these features were calculated on the bronchi and aerated lung regions separately, and used as independent hyper-curvature features.

The above DRF vectors, radiomic hyper-curvature features, and the combination thereof were subjected to a Cox proportional hazards model with an elastic-net penalty (hereafter called elastic-net Cox model) [5] for selecting and combining the elements of the vectors to build a Cox model that optimally predicts the survival of the patient. We used the concordance index (C-index) as a performance metric of the elastic-net Cox model. Bootstrapping with 750 replications was used to obtain an unbiased estimate of the C-index values for the DRFs, hyper-curvature features, and the combination thereof, and they were compared with an established clinical prognostic biomarker for ILD known as the gender, age, and physiology (GAP) index [2] using two-sided t test.

Results

Figure 1 shows the notched box plots of the C-index values obtained by application of the Cox model with the following combinations of features: (a) GAP model: C-index 78.5%, [95% confidence interval (CI) 70.3, 86.8]; (b) radiomic hyper-curvature features (RHC): C-index 86.1%, [CI 79.8, 92.4]; (c) DRF: 90.6% [CI 82.6, 98.7]; and (d) DRF + RHC: 92.1% [CI 84.7, 99.5]. The results indicate that the predictive performance of both the radiomic hyper-curvature features and DRFs are significantly higher than that of the GAP index ($p < 0.001$). Moreover, the combined features, DRF + RHC, outperform either one of these features ($p < 0.001$), indicating that the DRFs and radiomic hyper-curvature features are mutually complementary in their prediction performance.

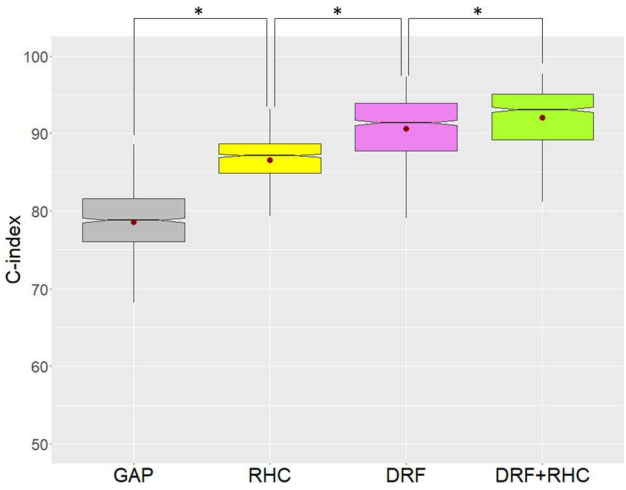

Fig. 1 Notched box plot of the C-index values obtained from the four different radiomic feature combinations. *Two-tailed t test: $P < 0.001$

Figure 2 shows the Kaplan–Meier survival curves of patients who were stratified to low- and high-risk groups of patients based on the combination of DRF and hyper-curvature features. Long-rank test