

Feasibility Study of MRI and Multimodality CT/MRI Radiomics for Lung Nodule Classification

Anthony E. A. Jatobá¹, Marcelo C. Oliveira¹,
Marcel Koenigkam-Santos², Paulo de Azevedo-Marques²

¹Programa de Pós-graduação em Informática (PPgI),
Instituto de Computação, Universidade Federal de Alagoas (UFAL)
Maceió, AL – Brasil

²Programa de Pós-graduação Interunidades em Bioengenharia,
Universidade de São Paulo (USP)
Ribeirão Preto, SP – Brasil

{aeaj,oliveiramc}@ic.ufal.br, {marcelk46,pmarques}@fmrp.usp.br

Abstract. Lung cancer is the most common and lethal form of cancer, and its early diagnosis is key to the patient's survival. CT is the reference imaging scan for lung cancer screening; however, it presents the drawback of exposing the patient to ionizing radiation. Recent studies have shown the relevance of MRI in lung nodules diagnosis. In this work, we aimed to evaluate whether radiomics features from MRI are well-suited for lung nodules characterization and if the combination of CT and MRI features can yield better results than the features from the individual modalities. For such, we segmented paired CT and MRI nodules from 33 lung nodules patients, extracted 89 radiomics features from each modality, and combined it into a multimodality feature set. Those features were then used for classifying the nodules into benign and malignant by a set of machine learning algorithms, assessing the AUC across 30 trials. Our results show that MRI radiomics features are suitable for characterizing lung lesions, yielding AUC values up to 17% higher than their CT counterparts, and shedding light on MRI as a viable image modality for decision support systems. Conversely, our multimodality approach did not improve performance compared to the single-modality models, suggesting that the direct combination of multimodality features might not be an adequate strategy for dealing with multimodality medical images.

Resumo. O câncer de pulmão é o tipo mais frequente e letal de câncer e o seu diagnóstico precoce é crucial para a sobrevivência do paciente. A tomografia computadorizada (TC) é o padrão-ouro para o rastreamento da doença; no entanto, apresenta a desvantagem de expor o paciente à radiação. Estudos recentes têm demonstrado o potencial da ressonância magnética (RM) no diagnóstico de nódulos pulmonares. Este trabalho busca avaliar a aplicação de características radiômicas de RM para a caracterização de nódulos pulmonares e se a combinação de atributos de TC e RM podem levar a melhores resultados que as modalidades individuais. Para tal, foram segmentados nódulos pulmonares em imagens de TC e RM de 33 pacientes com nódulos pulmonares; a partir de cada modalidade foram extraídos 89 características radiômicas, que foram

combinadas em um conjunto de características multimodalidade. Estas características foram usadas para classificar os nódulos entre benignos e malignos por meio de algoritmos de aprendizagem de máquina, calculando a AUC em 30 iterações. Os resultados indicam que características radiômicas de RM são adequadas para a caracterização de lesões pulmonares, com valores de AUC até 17% maiores que seus equivalentes em TC e evidenciando RM enquanto modalidade de imagem para sistemas de suporte à decisão. No entanto, a abordagem multimodalidade não apresentou ganhos em desempenho, sugerindo que a concatenação de características pode não ser uma estratégia adequada para lidar com imagens médicas multimodalidade.

1. Introduction

Lung cancer is the leading cause of cancer incidence and mortality worldwide, with 2.1 million new cases and 1.8 million deaths in 2018 [Wild et al. 2020]. The time of detection is a determinant factor for the prognosis; e.g., when a nodule is detected in its early stages, survival rates can reach up to 90%, in contrast to a mere 15% when diagnosed in its last stages [Blandin Knight et al. 2017]. Therefore, early diagnosis is a decisive factor for patients' treatment and survival [Siegel et al. 2018].

Early detection can be achieved by employing screening programs with computed tomography (CT). However, despite being the gold standard in lung cancer screening, CT still has shortcomings; since the exam requires a considerable radiation dose, performing periodic examinations can become undesirable because of the risks of radiation-induced cancers [Knight et al. 2017]. Low-dose computed tomography (LDCT) addresses this issue by requiring a significantly lower radiation dose, whose safety was backed up by several trials [Pastorino et al. 2012, Rampinelli et al. 2017]. Nevertheless, the reduced radiation dose from LDCT provides noisier images that can result in a high rate of false positives, leading to unnecessary invasive procedures [Li et al. 2019b].

In recent years, technical advancements in sequencing, scanners, and coils made magnetic resonance imaging (MRI) a viable modality for the management of patients with chest diseases such as lung cancer [Yi et al. 2008]. Besides not exposing the patient to radiation, MRI presents certain advantages over CT and LDCT, such as superior soft tissue contrast, allowing for better characterization of tumors [Yi et al. 2008, Beckett et al. 2015]. However, thoracic MRI is not as established as CT imaging; therefore, further clinical trials and development of protocols are necessary to determine its actual capacities. Nevertheless, preliminary works show the benefits of using MRI as a complement to the standard CT screening protocol [Sieren et al. 2010, Ohno et al. 2018].

Due to recent developments in hardware and software, multimodality medical imaging has been progressively applied in research and clinical practice [Wei et al. 2019]. The hypothesis behind multimodality imaging is that different modalities can provide complementary information of a disease, allowing for better characterization and support for decision making and treatment [Wei et al. 2019]. This area also holds potential for computer-aided detection and CADx research, with several works suggesting that the combination of different imaging modalities can bring better results in segmentation [Guo et al. 2019], therapy planning [Vaidya et al. 2012], and prognosis [Wei et al. 2019], when compared to single modality imaging.

Although there is a vast literature on lung cancer CT-based computer-aided diagnosis (CADx) systems, few studies have addressed the applicability of MRI for this type of cancer [Yang et al. 2018, Francisco et al. 2019]. Moreover, while significant progress has been made in assessing multimodal imaging on the diagnosis of several diseases, there is still a gap in the applicability of such combinations in regards to lung cancer. Lastly, CT/MRI is not a usual combination of imaging modalities for lung cancer assessment, making it beneficial to investigate the combination of these images' potential in CADx systems design.

This paper's main goal was to assess whether MRI radiomics features are well-suited for lung nodules characterization. As a secondary objective, we examined if the combination of CT/MRI features can yield better results than each separate modality.

The work's main contribution is showing that MRI radiomics features are suitable for lung nodules characterization. To the best of our knowledge, this was the first time a study investigated the effectiveness of MRI radiomics features in the classification of pulmonary nodules. Our results also showed that the direct combination of multimodality features might not be an adequate strategy when dealing with more than one medical image modality.

2. Related Works

Due to recent technological developments in hardware and software, multimodality imaging techniques have been increasingly applied in clinical practice and research. Previously, multimodality imaging application was mainly related to the combination of anatomical and functional images to improve diagnostic accuracy or target definition. More recently, the fusion of various images, such as CT, positron emission tomography (PET), and MRI, has become more prevalent, enabling new applications and better characterization of a disease [Vaidya et al. 2012, Vallières et al. 2015, Guo et al. 2019]. Applying machine learning methods across multimodality images has shown better results than single modality modeling for prognostic and prediction of clinical outcomes, holding great potential for precision medicine [Wei et al. 2019].

Vaidya *et al.* [Vaidya et al. 2012] used pre-treatment PET and MRI images from 27 patients diagnosed with non-small-cell lung carcinoma (NSCLC) to evaluate post-radiotherapy tumor progression in terms of local and loco-regional recurrence. Thirty-two handcrafted features were extracted from both PET and MRI, including statistical descriptors of each modality, total lesion glycolysis of PET images, intensity volume histogram (I_X and V_X features), and texture features using the co-occurrence matrix. The predictive value of these metrics was evaluated using Spearman's rank correlation coefficient (rs) and multivariate logistic regression. A two-parameter model using PET and MRI attributes yielded a gain of performance of 30% for loco-regional and 7% for local failure compared to single modality models, holding promise as an approach to allow for more individualized treatments.

Vallières *et al.* [Vallières et al. 2015] combined PET and MRI features for evaluation of lung metastasis risk in soft tissue sarcomas (STS). Nine non-texture and forty-one texture features were extracted from the tumor region of separate (PET, T1, and T2) and fused (PET/T1 and PET/T2) scans from 51 patients with STSs of the extremities. Image fusion was performed using the wavelet transform, and the influence of different extrac-

tion parameters on the predictive value of textures was investigated. The model consisted of a logistic regression classifier that used four texture features from the fused PET/T1 and PET/T2 scans, reaching an AUC of 0.984 ± 0.002 in bootstrapping validation. PET features presented a higher predictive value than MRI; however, the addition of MRI information to PET significantly improved performance.

Mu *et al.* [Mu et al. 2018] investigated PET and MRI images for the prediction of immunotherapy response in 64 NSCLC patients. The authors extracted 195 features from the original images and 1,235 features from images fused with multiple methodologies. The best model was a Support Vector Machine (SVM) classifier using 87 fused features and 13 single-modality features, yielding an AUC of 0.82, an improvement of 0.14 in AUC compared to only single-modality features.

Guo *et al.* [Guo et al. 2019] proposes an algorithmic architecture for supervised multimodal image analysis and categorizes feature fusion at Feature Level, Classifier Level, and Decision-Making Level. The authors designed a deep convolutional neural network system for image segmentation to contour lesions of STS using multimodal images of CT, MRI, and PET. The network trained with multiple modalities presented superior performance to networks trained with single modal images. According to the author, performing image fusion within the network layers is generally better than fusing the features at the network output.

Li *et al.* [Li et al. 2019a] proposes a deep learning method to fuse multimodality information for tumor segmentation in PET/MRI. The solution consists of a 3D fully convolutional network to produce a probability map for MRI segmentation, followed by a fuzzy variational model to incorporate the probability map and the PET intensity for an accurate multimodality tumor segmentation. The experimental results demonstrated that this model is suitable for small datasets and can outperform the existing deep learning-based multimodality segmentation methods, with a dice similarity index of 0.86 ± 0.05 .

Studies on multimodality imaging for lung cancer assessment are relatively recent and scarce. To the best of our knowledge, no works were developed regarding lung nodules classification in CT/MRI sequences. Because MRI is not widely used in clinical practice and its capacity is still being investigated, research is severely impaired by this lack of data. Thus, our work intends to fill those gaps in the knowledge by performing an investigation on how these image modalities can be integrated into a more precise lung cancer diagnosis support.

3. Material and Methods

In this section, we describe the steps taken in our methods. We present all details related to our lung nodule CT and MRI database in Section 3.1. First, we segmented the nodules using a semi-automatic algorithm (Section 3.2). Next, we extracted shape and texture radiomics features from every lesion in each imaging modality (MRI and CT) and combined the features into a fused dataset (Section 3.3). Lastly, we evaluated the performance of each set of features using 5-fold nested cross-validation, with the inner loop performing feature selection and hyperparameter optimization (Section 3.4).

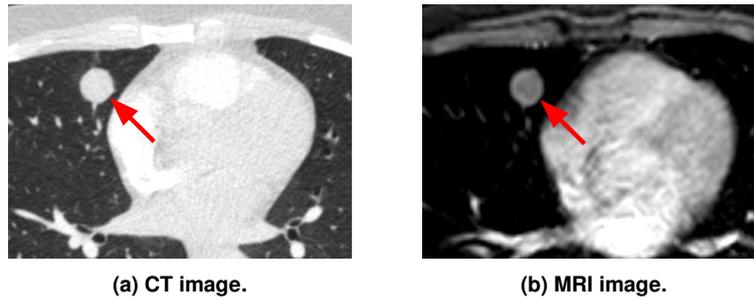


Figure 1. Lung nodule as seen in each imaging modality.

3.1. Database of Pulmonary Nodules in CT and MRI

We considered a cohort of 33 patients for this work. Our institutional research board approved this prospective study (HCRP process number: 3733/2017) with all patients' informed consent. All exams were anonymized to ensure patients' privacy.

We acquired CT imaging with a CT scanner (Philips, Big Bore) and MRI using a 1.5T device (Phillips, Achieva). The sequences were obtained with the patients in the supine position and with deep inspiration breath-hold procedure. The clinical chest MRI protocol included the T1 post-contrast (T1PC) sequence that approximates post-contrast CT images, with proper spatial and contrast resolution. Figure 1 shows a lung nodule as seen in CT (1a) and MRI (1b).

A senior radiologist pinpointed the location of the lesions on the CT and MRI sequences. The diagnosis was assessed after pathological confirmation of clinical treatment and stability. Our entire image database comprises 33 nodules with a size equal to or greater than 10mm, of which 21 were diagnosed as malignant and 12 as benign.

3.2. Nodules Segmentation

For every modality, each lesion was segmented using the semi-automatic segmentation algorithm FastGrowCut [Zhu et al. 2014]. This algorithm requires the radiologist to manually draw seed regions of the tissues to be segmented in each anatomical plane. These labels are then propagated based on principles from cellular automata to classify all the voxels as foreground or background.

To highlight the lung anatomical structures and better identify the tissues, we applied a greyscale lung windowing by setting the window in 1,400 and level in -500 Hounsfield unit (HU) in the CT images. For the MRI sequences, the radiologist set the level and width values to 800 and 2,000, respectively.

3.3. Feature extraction

Radiomics is the field that aims to convert digital medical images into high-dimensional data for improved decision support (e.g., diagnosis, prognosis assessment, therapy response) and precision medicine (e.g., therapy planning) by extracting a large number of handcrafted features from a volume of interest (VOI) [Gillies et al. 2016]. These features can capture characteristics of this VOI that otherwise would be difficult or impossible to be discerned even by an experienced professional [Hatt et al. 2017].

We used the segmentation to extract a series of radiomics features from each lesion using the open-source library pyradiomics (version 3.0.0) [Van Griethuysen et al. 2017]. At the time of this work, pyradiomics supported the following feature classes: First Order Statistics; Shape-based; gray level co-occurrence matrix (GLCM); gray level run length matrix (GLRLM); gray level size zone matrix (GLSZM); neighboring gray tone difference matrix (NGTDM); and gray level dependence matrix (GLDM).

An intrinsic limitation of MRI is that its images have arbitrary intensity units, i.e., the images' grey levels have no physiological meaning, as opposed to the CT HU scale. This characteristic makes comparisons difficult among different exams and makes image quantification with histogram-based features impractical. For this reason, we discarded first-order statistics features.

Shape-based features take into account the morphological characteristics of the nodule (e.g., nodules with spiculated borders are suspicious for malignancy, while well-defined round nodules are usually benign) [Ferreira et al. 2018]. The remaining features contain texture information, taking into account the local intensity-spatial distribution. Those features' performances are not affected by tumor position, orientation, size, and brightness [Wei et al. 2019]. After the extraction, we obtained a set of 89 features for each lesion in each imaging modality, divided into 14 shape-based, 24 GLCM, 16 GLRLM, 16 GLSZM, 5 NGTDM, and 14 GLDM features. We normalized each feature value using Min-Max scaling.

For our multimodality CT/MRI approach, we combined the single modalities features into a new set. According to the abstraction of fusion strategies proposed by Guo *et al.* [Guo et al. 2019], this approach is equivalent to performing fusion at the classifier level (using images of each modality as separate inputs to learn individual feature sets and combine them to support the learning of a multimodal classifier). Thus, the fused CT/MRI feature set contains all the features from each modality, adding up to 178 radiomic features.

3.4. Experiment Design

Feature Selection

Because of the inherent high dimensionality of radiomics feature sets, it is essential to perform feature selection [Wei et al. 2019]. Feature selection aims to obtain the optimal feature set relevant and non-redundant to the task, improving model performance and avoiding overfitting [Ribeiro et al. 2009]. We used a filter method by ranking the best features according to the ANOVA F-value statistic. For the decision tree and random forest classifier, we considered all features. For the remaining algorithms, we evaluated sets of 5, 10, 20, and 30 features within our hyperparameter optimization strategy.

Oversampling

Dealing with unbalanced classes may lead to models biased in favor of the majority class. That is the case with our dataset, as the ratio of positive and negative cases is almost two-fold (1:75:1). To deal with this problem, we performed oversampling through the synthetic minority over-sampling technique (SMOTE). SMOTE works by introducing

synthetic examples along with line segments that join k minority class nearest neighbors [Chawla et al. 2002]. We evaluated k values of 3 and 5 for each classifier.

Machine Learning Algorithms

We selected a set of machine learning algorithms to represent a wide variety of classifiers and evaluate how each one performs in the task. The algorithms considered are logistic regression (LR), k-nearest neighbors (KNN), SVM, decision tree (DT), random forest (RF), naive bayes (NB), and multi-layer perceptron (MLP) neural network. The open-source machine learning library scikit-learn provided the algorithms’ implementations (version 0.22.2) [Pedregosa et al. 2011]. Although deep learning models, such as convolutional neural networks and deep belief networks, are state of the art in lung nodule classification, the size of our dataset is insufficient to take advantage of those models [Parekh and Jacobs 2019]. In this case, radiomics is more suitable for addressing the problem of classification in the context of smaller image datasets, with the advantage of allowing the radiologist to interpret the role of each feature in the classifier’s output [Gillies et al. 2016].

Hyperparameter Optimization

Several machine learning algorithms can be parametrized by hyperparameters. These parameters can modify certain aspects of the learning algorithm and can have considerable effects on performance [Jatobá et al. 2020]. For this work, we performed hyperparameter optimization using grid-search with the area under the receiver AUC as the scoring metric [Claesen and De Moor 2015]. The search space explored is shown in Table 1.

Table 1. Search space used for Grid-Search.

Model	Parameters	Values
Logistic Regression	Penalty	{11, 12}
	C	{ 10^{-4} , 10^{-3} , ..., 10^0 }
KNN	K neighbours	{3, 5, 7}
	Weight	{niform, distance}
SVM	C	{ 10^0 , 10^1 , ..., 10^4 }
	Gamma	{ 10^{-4} , 10^{-3} , ..., 10^0 }
Decision Tree	Max Depth	{3, 4, 5, 6, 7, 8, 9, 10}
	Min Samples/Split	{2, 3, 4}
Random Forest	Number of estimators	{0, 100, 150}
	Max Depth	{2, 3, 4}
Naive Bayes	-	-
MLP	Hidden Layers	{(10), (20), (10, 5), (20, 10)}
-	Number of features	{5, 10, 20, 30}

Validation

Our evaluation methodology uses a set of metrics well-established in CADx systems design: AUC, sensitivity, and specificity. The AUC summarises the diagnostic ability of

binary classifiers, and it has been widely used in the medical literature [Fawcett 2006]. Hence, we used AUC for model comparison and optimization tasks. In addition, sensitivity and specificity analysis were used to better understand the models' capacity to classify positive and negative instances.

Our pipeline includes oversampling, feature selection, and hyperparameter optimization; it is necessary to perform a robust validation to ensure no data leakage occurs between our train and test folds, leading to unreliable results due to overfitting. Thus, we performed validation using 5-fold nested cross-validation [Cawley and Talbot 2010], performing the optimizations mentioned above only on the train folds. Finally, we repeated the experiment 30 times to obtain the average and deviation in performance for each metric.

4. Results and Discussion

Figure 2 presents the average performance of each classifier according to AUC for CT, MRI, and CT + MRI.

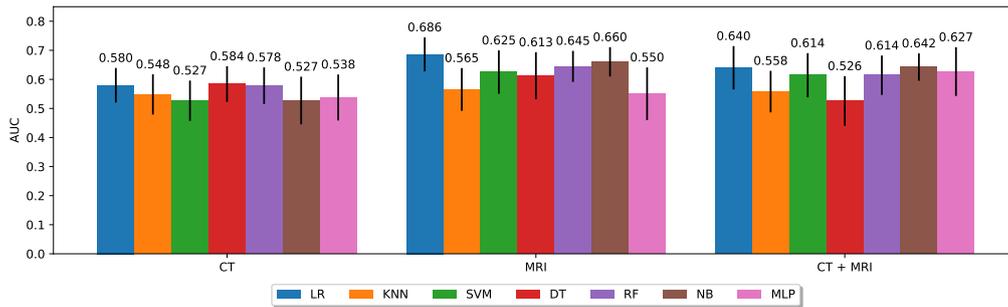


Figure 2. Average AUC for each classifier, by imaging modality.

4.1. MRI

Our results contain an intriguing finding, as the models trained with MRI radiomics presented superior AUC performance compared to those trained with features from CT imaging, the reference imaging for lung cancer screening. To assess the statistical significance of the classifiers' performance difference across the two datasets, we performed a Wilcoxon signed-rank test [Demšar 2006]. We verified the significance of this difference and rejected the null hypothesis with a confidence level of 5% for LR ($p = 2.85e-4$); SVM ($p = 1.11e-4$); RF ($p = 5.57e-4$); and NB ($p = 2.36e-6$). The confidence intervals in Figure 2 are a complementary tool to compare these results.

A conceivable explanation is that the superior soft-tissue contrast provided by MRI allowed for a better characterization of the tumors in terms of radiomics features and lead to a higher quality segmentation. According to the radiologist and as it can be observed in Figure 3, there is little contrast between the lesion and the mediastinum in the CT image (Figure 3a) compared to the MRI (Figure 3c). The consequence is that the segmentation in CT often invades the mediastinum structures (Figure 3b), while this does not happen in MRI (Figure 3d).

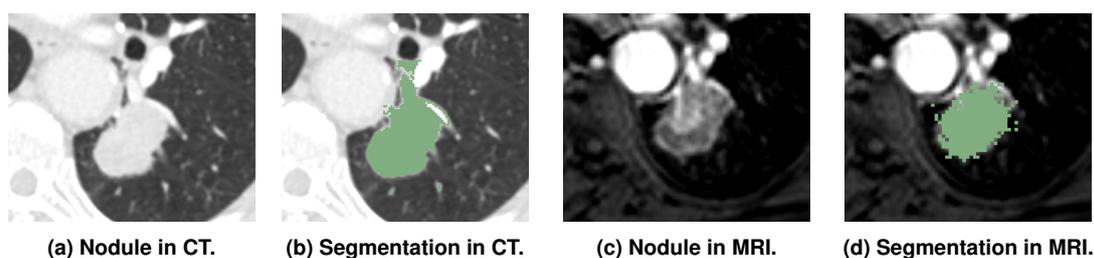


Figure 3. A figure with two subfigures

4.2. CT + MRI

We can also observe that the combination of CT and MRI features has not resulted in better classification performance, as MRI surpassed the combined models in every case, except for the MLP classifier. This classifier, however, could not outperform the best classifiers in the MRI set. This result may indicate that our feature selection approach was not suitable for combining multimodality radiomics features. The performance of the combined classifiers was expected to be at least equivalent to the classifiers in the individual modalities, as the set of combined features contains the features from both CT and MRI exams. Moreover, this result is in agreement with Wei *et al.*, who argue that the direct combination of features extracted separately might not make full use of the underlying biological correlation, making the information fusion strategy key to obtaining better models [Wei et al. 2019].

Going deeper into the models' behavior, we can notice that, in general, they have exhibited higher sensitivity than specificity (Figure 4); therefore, malignant nodules are more likely to be correctly classified. Although sensitivity is often favored over specificity in these tasks, the gap between those metrics is quite significant, suggesting that our oversampling strategy could not fully mitigate the imbalance in our data.

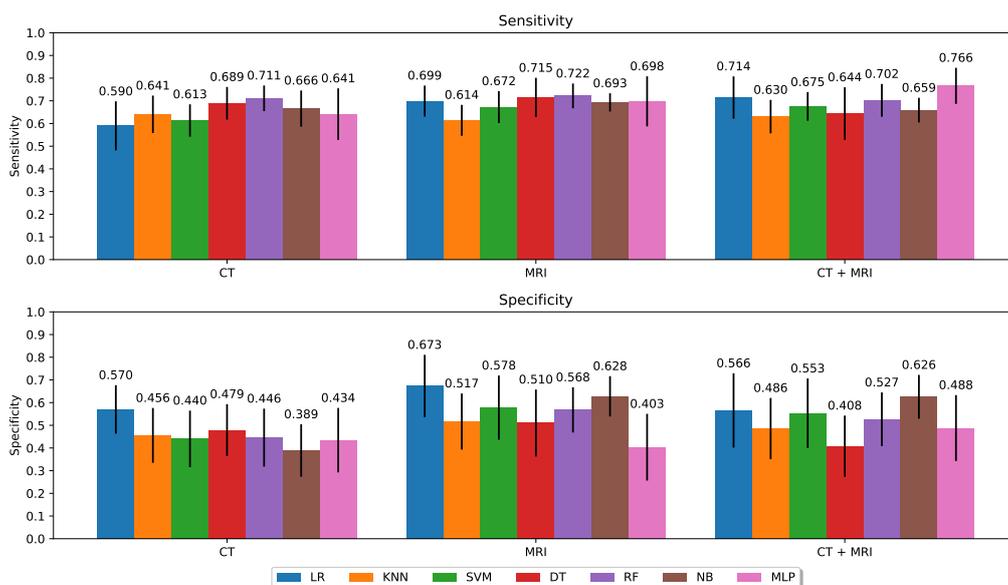


Figure 4. Average sensitivity and specificity for each classifier, by imaging modality.

5. Conclusion

This study aimed to evaluate the applicability of MRI and CT/MRI radiomics features for the characterization of lung nodules. Our results showed that MRI radiomics features can characterize lung nodules and support the development of predictive models, with AUC values up to 17% higher than their CT counterparts. This advantage over CT is exciting, as using MRI for lung cancer assessment can enable the investigation of physiological and functional aspects of a lung nodule instead of the primarily morphological information present on CT images. Moreover, the employment of MRI can mitigate problems such as radiation exposure and adverse reactions to contrast materials commonly used in CT.

On the other hand, our proposed multimodality method has not proven advantageous, with no gain in performance compared to the models using single modality MRI features. Acknowledging that CT is the gold-standard image modality for lung cancer diagnostic, we believe that a more sound investigation into multimodality medical imaging fusion techniques is needed. Furthermore, our results are promising and encourage more in-depth analysis into the potential of radiomics applied to MRI for lung cancer management, holding great promise for more accurate diagnostics, personalized treatment, and better outcomes.

References

- Beckett, K. R., Moriarity, A. K., and Langer, J. M. (2015). Safe use of contrast media: what the radiologist needs to know. *Radiographics*, 35(6):1738–1750.
- Blandin Knight, S., Crosbie, P. A., Balata, H., Chudziak, J., Hussell, T., and Dive, C. (2017). Progress and prospects of early detection in lung cancer. *Open biology*, 7(9):170070.
- Cawley, G. C. and Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, 11:2079–2107.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.
- Claesen, M. and De Moor, B. (2015). Hyperparameter search in machine learning. *arXiv preprint arXiv:1502.02127*.
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30.
- Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874.
- Ferreira, J. R., Oliveira, M. C., and de Azevedo-Marques, P. M. (2018). Characterization of pulmonary nodules based on features of margin sharpness and texture. *Journal of digital imaging*, 31(4):451–463.
- Francisco, V., Koenigkam-Santos, M., Wada, D. T., Junior, J. R. F., Fabro, A. T., Cipriano, F. E. G., Quatrina, S. G., and de Azevedo-Marques, P. M. (2019). Computer-aided diagnosis of lung cancer in magnetic resonance imaging exams. In *XXVI Brazilian Congress on Biomedical Engineering*, pages 121–127. Springer.

- Gillies, R. J., Kinahan, P. E., and Hricak, H. (2016). Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577.
- Guo, Z., Li, X., Huang, H., Guo, N., and Li, Q. (2019). Deep Learning-Based Image Segmentation on Multimodal Medical Imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2):162–169.
- Hatt, M., Tixier, F., Visvikis, D., and Le Rest, C. C. (2017). Radiomics in pet/ct: more than meets the eye? *Journal of Nuclear Medicine*, 58(3):365–366.
- Jatobá, A., Lima, L., Amorim, L., and Oliveira, M. (2020). Cnn hyperparameter optimization for pulmonary nodule classification. In *Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 25–36, Porto Alegre, RS, Brasil. SBC.
- Knight, S. B., Crosbie, P. A., Balata, H., Chudziak, J., Hussell, T., and Dive, C. (2017). Progress and prospects of early detection in lung cancer. *Open Biology*, 7(9).
- Li, L., Zhao, X., Lu, W., and Tan, S. (2019a). Deep learning for variational multimodality tumor segmentation in pet/ct. *Neurocomputing*.
- Li, S., Xu, P., Li, B., Chen, L., Zhou, Z., Hao, H., Duan, Y., Folkert, M., Ma, J., Huang, S., et al. (2019b). Predicting lung nodule malignancies by combining deep convolutional neural network and handcrafted features. *Physics in Medicine & Biology*, 64(17):175012.
- Mu, W., Qi, J., Lu, H., Schabath, M., Balagurunathan, Y., Tunali, I., and Gillies, R. J. (2018). Radiomic biomarkers from pet/ct multi-modality fusion images for the prediction of immunotherapy response in advanced non-small cell lung cancer patients. In *Medical Imaging 2018: Computer-Aided Diagnosis*, volume 10575, page 105753S. International Society for Optics and Photonics.
- Ohno, Y., Kauczor, H.-U., Hatabu, H., Seo, J. B., van Beek, E. J., and for Pulmonary Functional Imaging (IWPMFI), I. W. (2018). Mri for solitary pulmonary nodule and mass assessment: current state of the art. *Journal of Magnetic Resonance Imaging*, 47(6):1437–1458.
- Parekh, V. S. and Jacobs, M. A. (2019). Deep learning and radiomics in precision medicine. *Expert review of precision medicine and drug development*, 4(2):59–72.
- Pastorino, U., Rossi, M., Rosato, V., Marchianò, A., Sverzellati, N., Morosi, C., Fabbri, A., Galeone, C., Negri, E., Sozzi, G., et al. (2012). Annual or biennial ct screening versus observation in heavy smokers: 5-year results of the mild trial. *European Journal of Cancer Prevention*, 21(3):308–315.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Rampinelli, C., De Marco, P., Origgi, D., Maisonneuve, P., Casiraghi, M., Veronesi, G., Spaggiari, L., and Bellomi, M. (2017). Exposure to low dose computed tomography for lung cancer screening and risk of cancer: secondary analysis of trial data and risk-benefit analysis. *bmj*, 356:j347.

- Ribeiro, M. X., Balan, A. G., Felipe, J. C., Traina, A. J., and Traina, C. (2009). Mining statistical association rules to select the most relevant medical image features. In *Mining complex data*, pages 113–131. Springer.
- Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians*, 68(1):7–30.
- Sieren, J. C., Ohno, Y., Koyama, H., Sugimura, K., and McLennan, G. (2010). Recent technological and application developments in computed tomography and magnetic resonance imaging for improved pulmonary nodule detection and lung cancer staging. *Journal of Magnetic Resonance Imaging*, 32(6):1353–1369.
- Vaidya, M., Creach, K. M., Frye, J., Dehdashti, F., Bradley, J. D., and El Naqa, I. (2012). Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiotherapy and Oncology*, 102(2):239–245.
- Vallières, M., Freeman, C. R., Skamene, S. R., and El Naqa, I. (2015). A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Physics in Medicine and Biology*, 60(14):5471–5496.
- Van Griethuysen, J. J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G., Fillion-Robin, J.-C., Pieper, S., and Aerts, H. J. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21):e104–e107.
- Wei, L., Osman, S., Hatt, M., and El Naqa, I. (2019). Machine learning for radiomics-based multimodality and multiparametric modeling. *The quarterly journal of nuclear medicine and molecular imaging : official publication of the Italian Association of Nuclear Medicine (AIMN) [and] the International Association of Radiopharmacology (IAR), [and] Section of the Society of...*, 63(4):323–338.
- Wild, C. P., Weiderpass, E., and Stewart, B. W. (2020). *WORLD CANCER REPORT: cancer research for cancer development*. International Agency for Research on Cancer.
- Yang, Y., Feng, X., Chi, W., Li, Z., Duan, W., Liu, H., Liang, W., Wang, W., Chen, P., He, J., and Liu, B. (2018). Deep learning aided decision support for pulmonary nodules diagnosing: A review. *Journal of Thoracic Disease*, 10(Suppl 7):S867–S875.
- Yi, C. A., Shin, K. M., Lee, K. S., Kim, B.-T., Kim, H., Kwon, O. J., Choi, J. Y., and Chung, M. J. (2008). Non-small cell lung cancer staging: efficacy comparison of integrated pet/ct versus 3.0-t whole-body mr imaging. *Radiology*, 248(2):632–642.
- Zhu, L., Kolesov, I., Gao, Y., Kikinis, R., and Tannenbaum, A. (2014). An effective interactive medical image segmentation method using fast growcut. In *MICCAI workshop on interactive medical image computing*.