

Document downloaded from:

<http://hdl.handle.net/10251/108347>

This paper must be cited as:

Ortiz-Ramón, R.; Larroza-Santacruz, A.; Arana Fernandez De Moya, E.; Moratal, D. (2017). A radiomics evaluation of 2D and 3D MRI texture features to classify brain metastases from lung cancer and melanoma. Proceedings International Annual Conference of IEEE Engineering in Medicine and Biology Society. 493-496. doi:10.1109/EMBC.2017.8036869



The final publication is available at

<http://doi.org/10.1109/EMBC.2017.8036869>

Copyright IEEE Engineering in Medicine and Biology Society

Additional Information

A Radiomics Evaluation of 2D and 3D MRI Texture Features to Classify Brain Metastases from Lung Cancer and Melanoma

Rafael Ortiz-Ramón, Andrés Larroza, Estanislao Arana, and David Moratal, *Senior Member, IEEE*

Abstract— Brain metastases are occasionally detected before diagnosing their primary site of origin. In these cases, simple visual examination of medical images of the metastases is not enough to identify the primary cancer, so an extensive evaluation is needed. To avoid this procedure, a radiomics approach on magnetic resonance (MR) images of the metastatic lesions is proposed to classify two of the most frequent origins (lung cancer and melanoma). In this study, 50 T1-weighted MR images of brain metastases from 30 patients were analyzed: 27 of lung cancer and 23 of melanoma origin. A total of 43 statistical texture features were extracted from the segmented lesions in 2D and 3D. Five predictive models were evaluated using a nested cross-validation scheme. The best classification results were achieved using 3D texture features for all the models, obtaining an average AUC > 0.9 in all cases and an AUC = 0.947 ± 0.067 when using the best model (naïve Bayes).

I. INTRODUCTION

Brain metastases are more common than primary brain tumors and the survival of patients diagnosed with these lesions is estimated to be short, usually limited to months [1]. The exact incidence of brain metastases is unknown: some studies indicate that they occur in 9–17% of patients with cancer but these rates are thought to be higher [2], [3]. In adults, the primary tumors that metastasize more often to the brain are those originated in lung ($\geq 50\%$), breast (15–25%) and skin (melanoma) (5–20%) [1]. However, there is a small percentage of patients with brain metastases (2–14%) that are diagnosed with these lesions before detecting the location of the primary cancer [2]. Scarce and not recent literature concerning this group of patients is available and so their management remains unclear [4], [5]. For these patients, a fast, non-invasive and stable solution to identify the primary cancer would be required to determine the correct treatment.

Radiomics analysis in cancer studies has been proved to be a successful source of information to increase the precision in diagnosis, to evaluate the prognosis and to

predict treatment response [6]. This new radiomics concept is referred to the analysis of medical images by an exhaustive extraction of features from regions of these images and the corresponding data mining to create predictive models to help in the decision support [7], [8]. Radiomics comprises several fields, each of them extensively researched: medical imaging, segmentation, feature extraction and data mining. However, the focus of radiomics is the feature extraction step. To this end, texture analysis, which defines the quantification of gray-level patterns within the image, has been proved to be an excellent source of imaging biomarkers. Traditionally, texture analysis has been performed in 2D, but in the past years the biomedical community has made an effort to extend the 2D texture analysis techniques to the 3D space. This is a consequence of the improvement in the 3D data acquisition and the reaching of high spatial resolutions, which allow to capture tissue properties more accurately [9]. Volumetric texture analysis has been successfully applied in several studies involving cancerous brain lesions [10]–[12].

In this work, a radiomics approach to identify the primary site of origin of brain metastases by means of texture analysis was studied. We compared the discriminative power of 2D and 3D texture features extracted from contrast-enhanced T1-weighted magnetic resonance (MR) images of brain metastases from lung cancer and melanoma. Several predictive models were evaluated using this approach to study how different classifiers influence the results.

II. MATERIALS AND METHODS

A. Patients

A total of 30 patients (22 male and 8 female; age range: 24 – 73 years old, mean \pm standard deviation: 60.30 ± 11.45 years old) who were diagnosed with brain metastases between December 2013 and April 2016 were included in this study. Fifty initial metastatic lesions without previous treatment were found in these patients: 27 of lung and 23 of skin (melanoma) origin. Fig. 1 shows an example of both types of metastases.

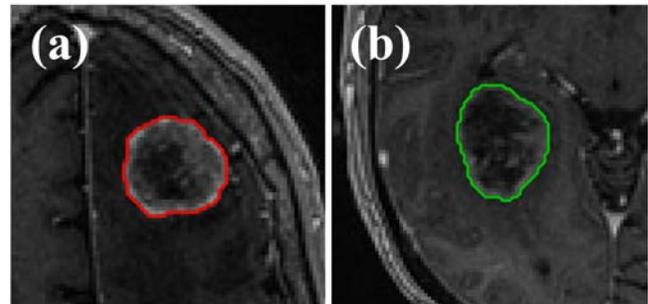


Figure 1. Examples of segmented brain metastases from lung cancer origin (a) and melanoma origin (b).

*Research supported in part by the Spanish Ministerio de Economía y Competitividad (MINECO) and FEDER funds under Grant BFU2015-64380-C2-2-R.

R. Ortiz-Ramón is with the Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, 46022 Valencia, Spain (e-mail: raorra@doctor.upv.es). He is supported by grant ACIF/2015/078 from the Conselleria d'Educació, Investigació, Cultura i Esport of the Valencian Community (Spain).

A. Larroza is with the Department of Medicine, Universitat de València, 46010 Valencia, Spain (e-mail: anlarro@gmail.com). He is supported by grant FPU12/01140 from the Spanish Ministerio de Educación, Cultura y Deporte (MECD).

E. Arana is with the Department of Radiology, Fundació Instituto Valenciano de Oncología, 46009 Valencia, Spain (e-mail: aranae@uv.es).

D. Moratal is with the Center for Biomaterials and Tissue Engineering, Universitat Politècnica de València, 46022 Valencia, Spain (phone: 96.387.70.00 ext. 88939; e-mail: dmoratal@eln.upv.es).

This retrospective, single-center study was approved by the Institutional Review Board of the Fundación Instituto Valenciano de Oncología, and all subjects provided written informed consent.

B. MRI Data

Imaging was performed using a 1.5T MRI scanner (Optima MR450w; GE Medical Systems, Milwaukee, WI, USA). MRI protocol included three-dimensional inversion recovery fast spoiled gradient-echo (IR-SPGR, BRAVO) T1-weighted images of the brain. Images were acquired without magnetization transfer, after intravenous administration of a single-dose of gadobenate dimeglumine (0.1 mmol/kg, MultiHance, Bracco; Milan, Italy) with a 6-min delay. Imaging parameters were: repetition time/echo time (TR/TE) of 8.5/2.2 ms; flip angle of 12°; slice thickness of 1.3 mm; pixel size of $0.98 \times 0.98 \text{ mm}^2$; and matrix size of 256×256 .

C. Image processing

Brain metastases were segmented in 2D and in 3D using a software tool developed in MATLAB (R2015b; The MathWorks Inc., Natick, MA, USA) specifically for this study. Each brain metastasis was manually segmented in 2D from the axial slice showing the largest lesion area. To segment the lesion in 3D, a semiautomatic method based on the Chan-Vese algorithm [13] was implemented. This method takes the previous manually segmented lesion area as a reference to segment the other slices containing metastatic tissue. The segmentation process was supervised by an expert radiologist with 20 years-experience in neuroradiology.

MR images were normalized using the $\mu \pm 3\sigma$ method to enhance the differences between classes, as proposed by Collewet *et al.* [14].

D. Feature extraction

To compute the features that describe the textures of the lesions, the MATLAB toolbox *Radiomics* implemented by Vallieres *et al.* [15] was used. This package allows extracting 43 texture features from 5 different statistical methods, both for 2D and for 3D texture analysis. In particular, 3 features were extracted from the intensity histogram (first-order

statistics) and the other 40 features were extracted from second-order statistical methods: 9 features derived from the gray-level co-occurrence matrix (GLCM), 13 from the gray-level run-length matrix (GLRLM), 13 from the gray-level size zone matrix (GLSZM) and 5 from the neighborhood gray-tone difference matrix (NGTDM). Information of each method and the corresponding features can be found in [15].

The second-order statistical features meet the criterion of rotation invariability. To this end, only one GLCM, GLRLM, GLSZM and NGTDM per lesion was computed. For 2D texture analysis, the neighboring properties of pixels in the 4 directions of the 2D space (0, 45, 90 and 135°) were averaged equally. For 3D texture analysis, the neighboring properties of voxels in the 13 directions of the 3D space were averaged differently to take into account discretization length differences [15]. To do this, the 3D regions need to be resampled to an isotropic voxel size; in this study, 3D regions were isotropically resampled to the in-plane resolution (voxel size = $0.98 \times 0.98 \times 0.98 \text{ mm}^3$) using cubic interpolation.

Prior to the computation of texture features, the intensity range of the metastatic regions was quantized to a lower number of gray levels (32) to improve the signal-to-noise ratio [16]. Finally, all texture features were standardized to zero mean and unit variance to avoid model computation being affected by the differences in the feature scales [17].

E. Model evaluation

Five different predictive models were studied to evaluate the discrimination power of the 2D and 3D texture features: naïve Bayes classifier (NB), k-nearest neighbors (k-NN), multilayer perceptron (MLP), random forests (RF) and support vector machine (SVM) with linear kernel. We chose five common classifiers from different predictive families to see which of them provides the best classification accuracy and to verify if there are noteworthy differences between 2D and 3D texture analysis using different approaches.

A nested cross-validation (CV) structure was used to evaluate the performance of each model (Fig. 2) without holding out some of the samples as an independent test set.

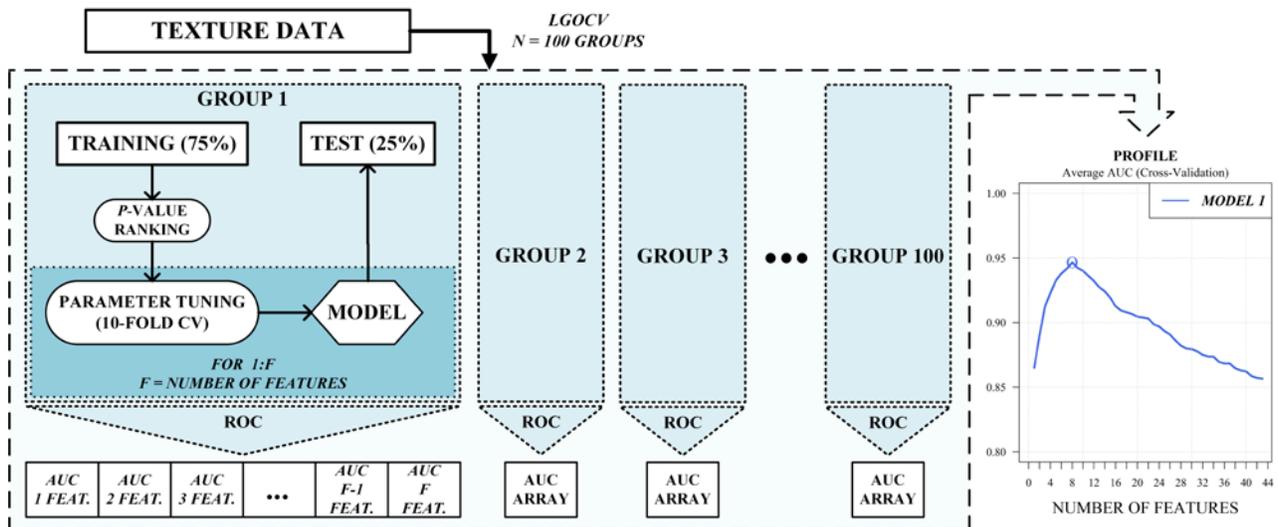


Figure 2. Structure of the nested CV process used to evaluate the different predictive models. Texture data extracted from the metastases were randomly divided into training and test sets $N = 100$ times to evaluate the model using different sets of samples and to obtain averaged results with low variance.

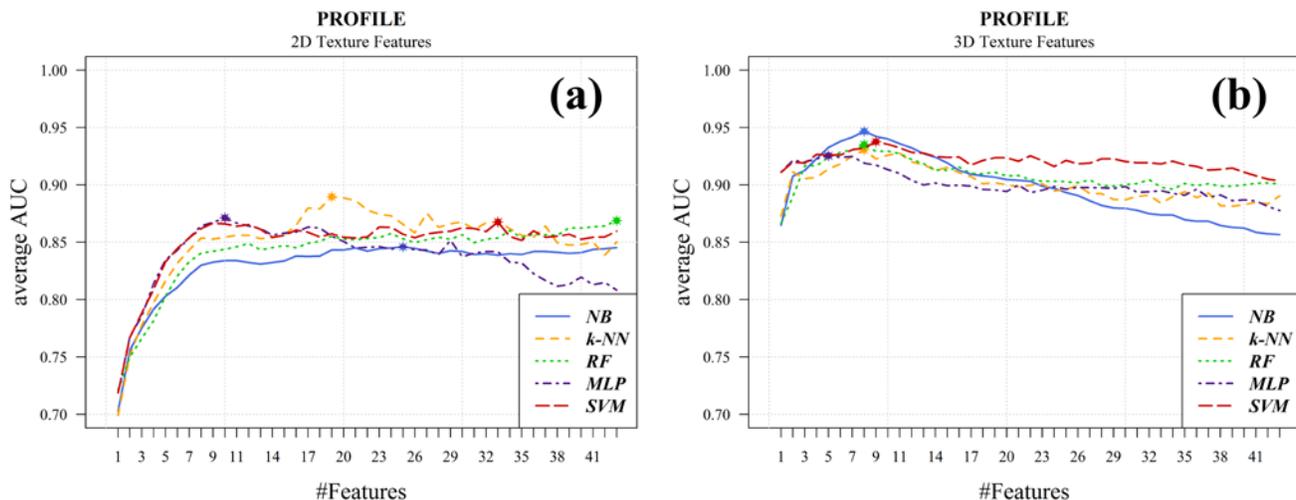


Figure 3. Profiles obtained after applying the nested CV process to the five models under analysis for 2D (a) and 3D (b) texture features. Both profiles show the average AUC obtained for each of the feature subsets.

We used this approach because the sample size of our dataset is relatively small and, in this situation, it is recommended to use every sample for model building. Proper estimates of model performance can be achieved using resampling methods when the number of samples is not large [17].

Leave-group-out CV (LGOVCV) was applied in the outer loop. This resampling method randomly divides the dataset into a training and a test set a total on N times, forming N groups. Each group is examined independently: the training set of a group is used to build the model and then this model is evaluated using the test set of the same group. At the end, the classification results provided by the estimates of all groups are averaged. In this study, a value of $N = 100$ groups was chosen to obtain results with low variance and consequently to decrease the uncertainty of the performance estimates [17]. In each group, 25% of the dataset was randomly selected as the test set and the remaining 75% was used as the training set. Model performance was evaluated using the area under the ROC curve (AUC) averaged over groups' estimates (mean \pm standard deviation).

The feature selection step was computed within the model-building process using the training set of each group. This process was not computed as an independent step to avoid overfitting, as mentioned by Ambroise and McLachlan [18]. A filter feature selection method based on the p -value provided by the Welch's t -test was employed to generate a ranking of the features with the most discriminative power. This method evaluates the statistical significance of each feature independently, without analyzing the relation between features and without involving any predictive model [17].

Parameter tuning was computed using the training set of each group by performing an inner 10-fold CV loop. This step was performed $F = 43$ times in each group estimate: the ranked features provided by the feature selection step were progressively added one by one from most to least important and then each feature subset was used to train the predictive model and to compute the AUC on the test samples of the same group. At the end, a total of $F = 43$ AUC values are provided in each group evaluation, one per each feature subset. The NB was evaluated without parameter tuning,

using a Gaussian kernel. The number of neighbors (k) in k -NN was selected from $k \in \{1, 3, 5, \dots, 21, 23\}$. The number of variables randomly sampled as candidates at each split ($mtry$) in RF was chosen from $mtry \in \{2, 3, 4, \dots, 19, 20\}$. The number of units in the hidden layer (l) of the MLP was selected from $l \in \{3, 5, 7, \dots, 23, 25\}$. SVM cost parameter (C) was chosen from $C \in \{2^{-4}, \dots, 2^0, \dots, 2^4\}$.

This model evaluation process was implemented with the Caret package [19] in R version 3.2.5 (R Development Core Team, Vienna, Austria). Model performance was analyzed separately for 2D and 3D features to study the differences.

III. RESULTS

Results show that classification with 3D texture features provides better accuracy than classification with 2D texture features. As it is illustrated in the profiles shown in Fig. 3, better average AUC is achieved when using 3D features (AUC > 0.9) instead of 2D features (AUC < 0.9) for all possible feature subsets and for all the models studied. Furthermore, as shown in Table I, the best AUC value in 3D texture analysis was obtained using fewer features ($\#Features \leq 9$) than in 2D texture analysis ($\#Features \geq 10$) for all predictive models. This is an important result because it indicates that these two types of lesions could be differentiated using only a few features, thus reducing the computation time.

TABLE I. COMPARISON BETWEEN MODEL PERFORMANCE RESULTS USING 2D AND 3D TEXTURE FEATURES

Model	2D texture features		3D texture features	
	#Features	AUC	#Features	AUC
NB	25	0.846 \pm 0.097	8	0.947 \pm 0.067
k-NN	19	0.890 \pm 0.085	8	0.930 \pm 0.069
RF	43	0.869 \pm 0.088	8	0.935 \pm 0.068
MLP	10	0.871 \pm 0.098	5	0.925 \pm 0.067
SVM	33	0.868 \pm 0.097	9	0.938 \pm 0.065

TABLE II. TOP TEN 3D TEXTURE FEATURES

Method	Feature	Average Rank	Average p -value
GLCM	Variance	1.22	$< 10^{-6}$
GLSZM	Low Gray-Level Zone Emphasis (LGZE)	2.05	$< 10^{-5}$
GLSZM	Small Zone Low Gray-Level Emphasis (SZLGE)	3.48	$< 10^{-5}$
GLRLM	Short Run Low Gray-Level Emphasis (SRLGE)	4.20	$< 10^{-5}$
GLRLM	Low Gray-Level Run Emphasis (LGRE)	5.53	0.00001
GLCM	Sum Average	5.78	0.00003
GLRLM	High Gray-Level Run Emphasis (HGRE)	6.50	0.00007
GLSZM	High Gray-Level Zone Emphasis (HGZE)	8.01	0.00020
GLRLM	Long Run Low Gray-Level Emphasis (LRLGE)	9.28	0.00048
GLSZM	Gray-Level Non-uniformity (GLN)	9.73	0.00075

Table I also shows that the classifier that yields the best AUC value for the 3D texture analysis is the NB (AUC = 0.947 ± 0.067). This result can be surprising because NB is the simplest of all the models studied. However, several studies related to this work also concluded that NB provides better accuracy results than other models [11], [20]. Anyhow, the five models produce a good classification accuracy, so it suggests that all models could be employed to discriminate metastases from lung cancer and melanoma.

Concerning the features with the most discriminative power, the nested CV scheme does not allow determining the exact ranking of features because the feature selection step is recomputed for every group. However, an average ranking was obtained. Table II shows the top ten 3D texture features used to evaluate the models: features derived from GLCM, GLRLM and GLSZM topped the ranking. Table II also shows the average p -value computed for these features: significant p -value ($p < 10^{-4}$) was obtained for these ten features, which gives an idea of their discriminative power.

Finally, it is relevant to mention that 3D features derived from the histogram and the NGTDM were not influential at all for classification, as none of the classifiers employed more than 9 features to achieve the best accuracy.

IV. CONCLUSION

In this study, a methodology based on a radiomics scheme was proposed to prove that brain metastases from lung cancer and melanoma can be differentiated by means of texture analysis. The results show that 3D texture features allow classifying both types of lesions more accurately than 2D texture features for all the models tested in this work. Furthermore, the highest AUC value was achieved using only 9 or less 3D texture features for all models. The next stage of this work would be to increase the number of samples to have the chance to create a final predictive model. It would also be interesting to include other primary sites of origin, like breast or kidney cancer.

REFERENCES

- [1] R. Soffietti *et al.*, "EFNS Guidelines on diagnosis and treatment of brain metastases: Report of an EFNS Task Force," *Eur. J. Neurol.*, vol. 13, no. 7, pp. 674–681, 2006.
- [2] L. Nayak, E. Q. Lee, and P. Y. Wen, "Epidemiology of brain metastases," *Curr. Oncol. Rep.*, vol. 14, no. 1, pp. 48–54, 2012.
- [3] I. T. Gavrilovic and J. B. Posner, "Brain metastases: Epidemiology and pathophysiology," *J. Neurooncol.*, vol. 75, no. 1, pp. 5–14, 2005.
- [4] R. Rudà, M. Borgognone, F. Benech, E. Vasario, and R. Soffietti, "Brain metastases from unknown primary tumour: a prospective study," *J. Neurol.*, vol. 248, no. 5, pp. 394–8, 2001.
- [5] S. Campos *et al.*, "Brain metastasis from an unknown primary, or primary brain tumour? A diagnostic dilemma," *Curr. Oncol.*, vol. 16, no. 1, pp. 62–6, Jan. 2009.
- [6] S. S. F. Yip and H. J. W. L. Aerts, "Applications and limitations of radiomics," *Phys. Med. Biol.*, vol. 61, no. 13, pp. R150–66, 2016.
- [7] V. Kumar *et al.*, "Radiomics: the process and the challenges," *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1234–1248, 2012.
- [8] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data.," *Radiology*, vol. 278, no. 2, pp. 563–77, Feb. 2016.
- [9] A. Depeursinge, A. Foncubierta-Rodriguez, D. Van De Ville, and H. Müller, "Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities," *Med. Image Anal.*, vol. 18, no. 1, pp. 176–196, 2014.
- [10] P. Georgiadis *et al.*, "Enhancing the discrimination accuracy between metastases, gliomas and meningiomas on brain MRI by volumetric textural features and ensemble pattern recognition methods," *Magn. Reson. Imaging*, vol. 27, no. 1, pp. 120–130, 2009.
- [11] A. E. Fetit, J. Novak, A. C. Peet, and T. N. Arvanitis, "Three-dimensional textural features of conventional MRI improve diagnostic classification of childhood brain tumours," *NMR Biomed.*, vol. 28, no. 9, pp. 1174–1184, 2015.
- [12] D. Mahmoud-Ghoneim, G. Toussaint, J. M. Constans, and J. D. De Certaines, "Three dimensional texture analysis in MRI: A preliminary evaluation in gliomas," *Magn. Reson. Imaging*, vol. 21, no. 9, pp. 983–987, 2003.
- [13] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, 2001.
- [14] G. Collewet, M. Strzelecki, and F. Mariette, "Influence of MRI acquisition protocols and image intensity normalization methods on texture classification," *Magn. Reson. Imaging*, vol. 22, no. 1, pp. 81–91, 2004.
- [15] M. Vallières, C. R. Freeman, S. R. Skamene, and I. El Naqa, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Phys. Med. Biol.*, vol. 60, pp. 5471–96, 2015.
- [16] P. Gibbs and L. W. Turnbull, "Textural analysis of contrast-enhanced MR images of the breast," *Magn. Reson. Med.*, vol. 50, no. 1, pp. 92–98, Jul. 2003.
- [17] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, 1st ed. Springer-Verlag New York, 2013.
- [18] C. Ambrose and G. J. McLachlan, "Selection bias in gene extraction on the basis of microarray gene-expression data.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 10, pp. 6562–6566, 2002.
- [19] M. Kuhn, "Building Predictive Models in R Using the caret Package," *J. Stat. Softw.*, vol. 28, no. 5, pp. 1–26, 2008.
- [20] J. Huang, J. Lu, and C. X. Ling, "Comparing naive Bayes, decision trees, and SVM with AUC and accuracy," in *Third IEEE International Conference on Data Mining (ICDM'03)*, 2003, pp. 553–556.