



**HAL**  
open science

## Improved 18-FDG PET/CT diagnosis of multiple myeloma diffuse disease by radiomics analysis

Charles Mesguich, Elif Hindie, Baudouin Denis de Senneville, Ghoufrane Tlili, Jean-Baptiste Pinaquy, Gerald Marit, Olivier Saut

► **To cite this version:**

Charles Mesguich, Elif Hindie, Baudouin Denis de Senneville, Ghoufrane Tlili, Jean-Baptiste Pinaquy, et al.. Improved 18-FDG PET/CT diagnosis of multiple myeloma diffuse disease by radiomics analysis. Nuclear Medicine Communications, 2021, 42 (10), pp.1135-1143. 10.1097/MNM.0000000000001437 . hal-03428536

**HAL Id: hal-03428536**

**<https://inria.hal.science/hal-03428536>**

Submitted on 18 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Improved 18-FDG PET/CT Diagnosis of Multiple Myeloma Diffuse Disease  
by Radiomics Analysis**

Charles Mesguich<sup>123</sup>, Elif Hindie<sup>1</sup>, Baudouin Denis de Senneville<sup>3</sup>, Ghoufrane Tlili<sup>1</sup>, Jean-Baptiste Pinaquy<sup>1</sup>, Gerald Marit<sup>2</sup>, Olivier Saut<sup>3</sup>

1. CHU Bordeaux, Nuclear Medicine department, F-33000 Bordeaux, France
2. INSERM U1035, University of Bordeaux, Bordeaux, France
3. University of Bordeaux, IMB, UMR CNRS 5251, INRIA Project team Monc, Talence, France

*Corresponding author: Dr. Charles Mesguich*

*Service de Médecine Nucléaire*

*Hôpital Haut-Lévêque*

*Avenue Magellan*

*33604 Pessac, France*

*Tel : +33-5-57656335 ; Fax : +33-5-57656839*

*E-mail : [charles.mesguich@chu-bordeaux.fr](mailto:charles.mesguich@chu-bordeaux.fr)*

*Running title: PET/CT Radiomics for myeloma diffuse disease*

*Financial support: Philips Healthcare*

*Word count: 4436*

## **Abstract**

In Multiple Myeloma (MM) patients, diffuse infiltration of bone marrow can be diagnosed on MRI and is associated with poorer prognosis. On 18-FDG PET/CT, the other important imaging modality in MM, the diagnosis of diffuse disease by visual analysis can be challenging. Radiomics allows the extraction of large amount of data from images to individualize disease specific diagnostic or prognostic patterns. We aimed to develop a radiomics-based model derived from PET and CT images, that could improve the diagnosis of multiple myeloma diffuse disease on 18-FDG PET/CT.

**Methods** We prospectively performed PET/CT and Whole-Body MRI in 30 MM patients at initial diagnosis. MRI was the standard of reference for diffuse disease assessment. Twenty patients were assigned to the training set and 10 to the independent test set. Visual analysis of PET/CT was done by two nuclear medicine physicians separately, then by consensual reading. Spine volumes on CT and PET were automatically segmented and a total of 174 IBSI-compliant radiomics features from both volumes were extracted. Selection of best features in the training set was performed with Random Forest features importance combined with correlation analysis. Machine-learning algorithms were then trained on the selected features with cross-validation. Finally, the model was evaluated on an independent test set.

**Results** Out of the 30 patients, 18 had established diffuse disease on MRI. After consensus of PET/CT images, the sensitivity, specificity and accuracy of visual analysis was 67%, 75% and 70%, respectively, with a moderate kappa coefficient of agreement of 0.6. Five radiomics features were selected. On the training set, Random Forest classifier reached the highest mean accuracy of 0.91 (95% CI 0.90-0.92) with an AUC of 0.90 (95% CI 0.89-0.91) for diffuse disease diagnosis. On the independent test set, the model achieved an accuracy of 80%.

**Conclusions** Radiomics analysis of 18-FDG PET/CT images with machine-learning overcame the limitations of visual analysis, providing a highly accurate and more reliable diagnosis of diffuse bone marrow infiltration in multiple myeloma patients.

## **Introduction**

Multiple Myeloma (MM) is characterized by the proliferation of clonal plasma cells in the bone marrow. It is the second most common hematological malignancy. 18-FDG PET/CT and MRI have progressively replaced conventional skeletal survey for MM initial work-up, as these modalities more extensively evaluate tumor burden (1). 18-FDG PET/CT is superior to MRI for treatment response assessment, allowing to monitor tumor metabolic activity with higher specificity (2,3). It is able to detect medullary and extra-medullary focal lesions, that both carry a significant prognostic value. However, MM diffuse disease can be difficult to diagnose on 18-FDG PET/CT. Patients with diffuse disease belong to a high-risk category, with frequent association with high-risk cytogenetics features (4,5). While MRI is the modality of choice for the diagnosis of diffuse disease (6,7), the lower spatial resolution of 18-FDG PET/CT could make it less sensitive. Additionally, anemia or inflammation may increase bone marrow uptake, mimicking diffuse disease, potentially altering PET/CT specificity. Different criteria for the diagnosis of diffuse disease on 18-FDG PET/CT have been proposed (8,9). These criteria are based on semi-quantitative analysis of bone marrow uptake, but their diagnostic performance has not yet been evaluated. Additionally, such approach might be limited by not considering the heterogeneity of bone marrow uptake, that a visual approach could appreciate but not objectively quantify.

Radiomics involves the extraction of a large number of quantitative features from medical images, providing image-based signatures that can help for disease diagnosis, prognostication or treatment response assessment (10-12). Indices extracted from the histogram of voxel values and texture features are able to quantify intratumoral heterogeneity. We hypothesized that a

quantification of bone marrow heterogeneity through Radiomics analysis would help improving the diagnosis of diffuse disease on 18-FDG PET/CT in multiple myeloma patients.

We aimed to develop a radiomics-based model derived from PET and CT images, that could improve the diagnosis of multiple myeloma diffuse disease on 18-FDG PET/CT.

## **Material and Methods**

### *Patients*

Between February 2017 and February 2019, patients with newly diagnosed MM eligible for autologous stem cell transplantation were prospectively enrolled in this study. The diagnosis of MM was based on the IMWG criteria (13). The inclusion criteria included informed consent to undergo PET/CT and Whole-Body diffusion-weighted MRI at diagnosis. Exclusion criteria were corticosteroid therapy during the 2 weeks prior to imaging or having already started chemotherapy. The Ethics Committee of the District of Bordeaux approved this research.

### *<sup>18</sup>F-FDG PET/CT*

PET/CT examinations were performed either on a GE-710-16S PET-CT instrument (GE Healthcare, Chicago, IL, USA) or on a Philips VEREOS PET/CT instrument (Philips Healthcare, Eindhoven, The Netherlands) with an injected <sup>18</sup>F-FDG activity of 3 MBq/kg. Patients fasted for 6 hours before <sup>18</sup>F-FDG injection. PET/CT was performed from the vertex to the knees, with an average total scan time of 20 minutes. Image findings were reported to clinicians for staging purposes. For the purpose of the present radiomics' analysis, and following data anonymization, PET/CT results were reviewed on a dedicated workstation (Advantage Workstation; GE Healthcare, Chicago, IL) by two experienced nuclear medicine physicians (JB and GT) blinded to the MRI, clinical and biological data. Visual assessment of bone marrow was performed twice by each reader with sufficient time elapsed between reading

sessions for intra-observer and inter-observer variability assessment. Visual interpretation was based on the uptake intensity (bone marrow uptake > liver uptake) and heterogeneity of bone marrow uptake on PET/CT images (3). CT was performed with a tube voltage of 120 Kv. CT matrix size was  $512^2$  with a voxel size of  $1 \times 1 \times 2$  mm (Philips) and  $1 \times 1 \times 2.5$  mm (GE). PET matrix size was  $288^2$  (Philips)  $256^2$  (GE) with a voxel size of  $2 \times 2 \times 2$  mm (Philips) and  $3 \times 3 \times 3$  mm (GE). PET images were reconstructed with an iterative algorithm: 3 iterations, 5 subsets; 2 mm full width half maximum (Philips); 2 iterations, 24 subsets; 6.4 mm full width half maximum (GE).

#### *Whole-body diffusion-weighted MRI*

Whole-body diffusion-weighted MRI (WB-MRI) was the standard of reference for the diagnosis of diffuse disease. Examinations were performed using a 1.5 Tesla device (Aera; Siemens Healthineers, Erlangen, Germany). The following imaging protocol was applied for an average total scan time of 45 minutes. First, we obtained coronal T1-weighted turbo spin echo sequences at the pelvis (repetition time [TR] 785 ms; echo time [TE] 10 ms), coronal T2-weighted short-term inversion recovery sequences at the pelvis (TR 13850 ms; TE 90 ms), sagittal T1-weighted turbo spin echo sequences at the spine (TR 453 ms; TE 12 ms), and sagittal T2-weighted short-term inversion recovery sequences at the spine (TR 7450 ms; TE 62 ms). Second, axial diffusion-weighted (DW) sequences were acquired from the vertex to the knees seven to nine stacks with  $b$  values of 50 and 800  $s/mm^2$ . Each stack was composed of 50 slices of 5 mm thickness (TR 7959 ms, TE 61 ms, inversion time 180 ms).

After anonymization, MRI results were reviewed by one experienced radiologist blinded to the PET/CT, clinical and biological data. The definition of diffuse disease was adapted from previously published criteria (14), based on measurements performed on the lumbar spine, as follows: a vertebral body/vertebral disk signal <1.3 on T1-weighted images, a vertebral body/psoas muscle signal >2 on T2-weighted images, and a vertebral body/kidney signal >1 on

DW images. Mild infiltration was defined as a pathological finding only on T2-weighted sequences or on DW sequences. Moderate infiltration was defined as a pathological finding on both T2-weighted and DW images. Severe infiltration was defined as a pathological finding on T2-weighted, DW and T1-weighted images.

### *Feature extraction*

The whole spine volume of interest (VOI) of a single patient was manually segmented on a CT template scan using Horos software (v3.3.5, Nimble Co LLC d/b/a Purview Annapolis, MD USA) by an experienced nuclear medicine physician. Subsequently, we employed a 3D deformable image registration (DIR) algorithm to propagate the obtained VOI on the CT-scan of each of the 30 patients. The employed DIR algorithm was designed to provide voxel-wise 3D motion estimates, in order to cope with inter-patients' anatomical variabilities. We employed the EVOlution solution which was initially proposed for mono- and multi-modal DIR (15,16). A visual inspection of propagated delineations was done, and local manual corrections were performed when needed. Extracted Region of Interests (ROIs) corresponding to focal lesions were then subtracted from the CT and PET VOI of the spine. This allowed to analyze the whole bone marrow signal of the spine without focal lesions, in order to properly individualize diffuse disease.

Because two different PET/CT instruments were used for this study, PET VOI were resampled to a voxel size of 2mmx2mmx2mm and CT VOI were resampled to a voxel size of 1x1x2mm. PET images were normalized between patients by using the SUV formula. Radiomics features from VOI of CT and PET were extracted by using the Pyradiomics package (17). Most of the features extracted comply with the features definitions described by the Imaging Biomarker Standardisation Initiative (IBSI) (18). For the purpose of features extraction on CT, we used a fixed bin width of 10 HU between -1000 and 3000 HU in line with the absolute resampling

method (19). For PET SUV, we used a fixed bin width of 0.3 between an intensity of 0 and 20, following the absolute resampling method (19). Conventional PET metrics were extracted: SUVmax, SUVmean, SUVmin and Metabolic Tumor Volume. The following extracted radiomics features were: first order statistics (n=19), second order features derived from Gray Level Cooccurrence matrix (n=20), third order features derived from Gray Level Run Length matrix (n=16), from Gray Level Size Zone matrix (n=16), from Gray Level Dependent matrix (n=12) and from Neighboring Gray Tone Difference matrix (n=4). Radiomics features are listed in the supplementary materials.

### *Modelling*

For each patient, a total number of 174 Radiomics features were obtained from CT (N=87) and PET images (N=87). Data analysis was performed using the SciKit Learn library (20). Patients were randomly assigned to a training set (2/3) or a test set (1/3). Features were pre-selected on the training set by Random Forest tree importance with a cut-off  $>0.01$  (21). A correlation matrix of the pre-selected variables was performed (figure 1) and features with high correlation were discarded: a total of five features were kept in the model. Machine-learning (ML) methods were trained to find the best model, using the selected features. Different ML classifiers from the Scikit Learn library were considered: logistic regression, single vector machine, k-nearest neighbors, gaussian naïve bayesian classifiers and random forests. Classifiers hyperparameters were optimized by cross-validated grid-search. Each model metrics were then evaluated using a 5-fold shuffled cross validation, which was iterated 100 times with a different initialization seed. The following average test metrics were calculated: recall (sensitivity), specificity, precision (positive predictive value), negative predictive value, F1-score [ $F1=2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$ ], accuracy and area under the curve after ROC analysis. The Radiomics flowchart is shown in figure 2.



## Results

Thirty consecutive patients were prospectively included in the analysis. The characteristics of the patients are summarized in Table 1. PET/CT and WB-MRI were performed in indeterminate order depending on scheduling opportunities. The median time between PET/CT and WB-MRI was 9 days (range, 1-23 days). The median glucose level in patients proceeding to PET/CT was 0.97 g/L (range, 0.78–1.37 g/L). The median interval between <sup>18</sup>F-FDG injection and scanning was 65.5 minutes (range, 45–70 minutes).

Eighteen patients (18/30) had diffuse bone marrow disease of the spine on WB-MRI which was classified as mild (n=1), moderate (n=8) or severe (n=9). The training set was composed of 13 patients with diffuse disease and 7 patients without diffuse disease. The accuracies, sensitivity and specificity of first and second PET/CT readers for detecting diffuse disease were 55%, 62%, 43% and 65%, 62%, 71%, on this training set respectively. The intra-observer variability was evaluated: the first reader had a Cohen's kappa coefficient of agreement of 0.71 ( $p < 0.001$ ) and the second reader 0.80 ( $p < 0.001$ ), which corresponds to a substantial agreement. After consensus reading, the combined reading accuracy reached 70% and the AUC of the ROC curve was 0.71 (CI 95% 0.54-0.88). Kappa coefficient of agreement between the two readers was 0.60 ( $P = 0.006$ ), which corresponds to a moderate agreement. Detailed diagnostic performances of readers are provided in Table 2.

Five radiomics features on PET-CT were selected in our final model. Two extracted from CT images were derived from the gray level co-occurrence matrix: Contrast and Joint Energy also defined as Angular Moment by IBSI. Three were extracted from PET images and were derived from the neighboring gray tone difference matrix: Busyness, Coarseness and Strength. There was no significant correlation between PET features ( $P > 0.05$ ) and CT features ( $P > 0.05$ ) as shown by the correlation matrix in figure 1. These parameters were also not correlated to their corresponding volume: Busyness -  $R = 0.14$  (95% CI -0.32-0.55;  $P = 0.55$ ); Coarseness -  $R =$

0.25 (95% CI -0.21-0.62;  $P=0.29$ ) ; Strength –  $R=0.18$  (95% CI -0.58-0.29;  $P=0.45$ ) ; Joint Energy -  $R = -0.07$  (95% CI -0.50-0.38;  $P=0.76$ ) ; Contrast –  $R= -0.39$  (95% CI -0.71-0.07 ;  $P=0.09$ ). There was no correlation between Coarseness and SUVmax ( $R=-0.18$ ; 95% CI -0.58-0.28 ;  $P=0.43$ ), no high correlation between Busyness and SUVmax ( $R= -0.56$ ; 95% CI -0.80-0.15;  $P=0.01$ ) and a high correlation between Strength and SUVmax ( $R=0.92$  ; 95% CI 0.80-0.97;  $P<0.001$ ). Performance of each classifier after 5-fold cross-validation with the training set using these five features are provided in Table 3. Random forest classifier reached the highest mean accuracy of 0.91 (95% CI 0.90-0.92) with an AUC of 0.90 (95% CI 0.89-0.91). ROC curves of the different classifiers are showed in figure 3. The independent test set was balanced with 5 patients with diffuse disease and 5 patients without diffuse disease. Out of the different classifiers evaluated, the trained Random Forest classifier reached the highest sensitivity on the independent test set with an accuracy of 80%. Gaussian Naïve-Bayes also showed similar accuracy in the test set. Detailed performance is provided in Table 3. Examples of diagnostic errors made by visual analysis with correct classification by texture analysis are shown in Figures 4 and 5. The most frequent errors made after texture analysis were similar between all the algorithms. The patient for whom a systematic error was made had a severe diffuse disease. Visual analysis for this patient was discordant between the two readers.

## **Discussion**

We aimed at developing a reliable and automated method based on computational imaging analysis of PET and CT images in order to accurately diagnose MM diffuse infiltrative disease at initial diagnosis. Currently, there are no recommended criteria for the diagnosis of diffuse disease on 18FDG PET-CT imaging. Authors have suggested using either the spleen or the liver uptake as a diagnostic cut-off, mainly based on an empirical approach (8,9). Therefore, current interpretation of PET/CT mainly relies on a visual approach. In our study, when PET/CT

images were read visually, accuracies of both readers were moderate. Agreement between readers was moderate and there was some intra-observer variation at two different readings. Moreover, accuracy only slightly improved after consensual reading (0.7) (Table-2). This reflects the limitations of a visual approach for the diagnosis of diffuse MM disease. The requirement for quantitative and/or standardized PET/CT interpretation criteria to be used in various settings of multiple myeloma patients have been pointed out (22).

We found that a radiomic signature based on five different features extracted from PET and CT images was the most accurate for the diagnosis of diffuse disease. The risk of overfitting our model on the training set was minimized by filtering features and operating cross-validation (23). After a 5-fold shuffled cross-validation iterated 100 times with a different initialization seed our model kept high performances, that were superior to visual reading. Furthermore, the model was stable after testing on an independent test set. In a post-hoc analysis, our model was also stable when randomizing 100 times the training and testing set distribution (training set: AUC=0.86;95% CI 0.84-0.88; testing set: accuracy = 0.80; 95%CI = 0.78-0.82). Interestingly, these five biomarkers were extracted from both CT and PET, highlighting the importance of both imaging modalities for the diagnosis of diffuse disease.

The non-correlation of these five parameters between themselves reflects that they represent different information. It has been demonstrated that there is a correlation between Radiomics parameters and volume but that this correlation decreases as the volume increases (10). This explains why our selected features presented no significant correlation with volume, as a very large volume was extracted for each patient (median Volume = 660682cm<sup>3</sup>). Although a high correlation was found between Strength and SUVmax, the two other selected PET features had none to moderate correlation to SUVmax, suggesting that the combination of the 3 PET features provides sufficient additional information as compared to SUVmax.

Two studies have shown that texture parameters have a valuable diagnostic value for the diagnosis of bone marrow infiltration in diffuse large B-cell lymphoma and acute leukemia respectively (24,25). As the diseases explored are different, the texture parameters that were found relevant in these studies are different from our model. The methodology used was different in the study by Aides and colleagues. In this study, a one-feature model was selected with univariate analysis (24). Skewness, a first order parameter, was selected as the best feature with a ROC AUC of 0.821 (24). Interestingly, in our study, this parameter also had a significant value with a ROC AUC of 0.78 and an accuracy of 0.70 on the training set. However, these results remained inferior to what we obtained with our model, and the single feature performed poorly in the test set (accuracy = 0.40). Busyness was the best feature after univariate analysis on the training set. However, using only Busyness as a predictor reached unsatisfying results on the test set with an accuracy of 0.67, which plead in favor of our model.

Improving PET/CT diagnosis of MM diffuse infiltrative disease is crucial for the following reasons: 1) 18-FDG PET/CT is the only imaging modality that has been proved as prognostic in the evaluation of MM response to treatment (2,3). Therefore, in order to properly assess disease response to treatment, it is necessary to have the most comprehensive PET/CT imaging study at baseline, being able to more accurately diagnose diffuse disease. 2) Although the diagnosis of diffuse disease is not a criteria that will trigger the start of a treatment, it is known that smoldering myeloma presenting with only diffuse infiltration are at high risk of progression to multiple myeloma (26). It is recommended that these patients be closely monitored with follow-up imaging every 3 to 6 months (13). 3) Diffuse disease intensity is not a simple reflect of bone marrow plasma cells percentage on the single BMB. Indeed, BMB cannot be a surrogate of imaging for diffuse disease assessment, probably partly because of the heterogeneity of plasma cells distribution throughout the entire bone marrow (27). This is the reason why we chose Whole-Body diffusion MRI as the reference standard in our study. This highlights the

pivotal role of MM whole-body imaging at diagnosis, including PET/CT, in order to improve MM patient risk stratification. We here provide a tool that improves diffuse disease assessment by FDG-PET contributing to make of FDG-PET the most comprehensive imaging procedure in MM, being able to accurately diagnose focal lesions, extra-medullary lesions and diffuse disease as well.

Our study had some limitations. First, although this study was conducted prospectively, the population of our study was small. Second, whether combination of radiomics features with some biological features might further improve prediction also deserves investigation on a larger patient sample. Third, although we used two different PET/CT systems, this study was monocentric, and external validation will be needed. Nevertheless, we tried to build a robust and standardized algorithm that comply with IBSI guidelines. As our final model relies identified imaging features extracted with an open-source package, we hope it will be further tested in studies from other centers.

## **Conclusion**

We elaborated a bone marrow radiomics-based model that allows to more accurately diagnose MM diffuse infiltrative disease on 18-FDG PET/CT. This reinforces 18-FDG PET/CT position as a one stop-shop procedure for assessing MM disease extent and being a baseline reference for subsequent therapeutic evaluation. Our results need validation in a larger and multicentric population.

## **Acknowledgments**

We would like to thank Philips Healthsystems for their intellectual contribution and financial support throughout this study.

## **Disclosure**

CM has received research support from Philips Healthsystems. No potential conflicts of interest relevant to this article exist.

### **Key Points**

**QUESTION:** Can Radiomics analysis of 18-FDG PET/CT images improve its performance for the diagnosis of diffuse disease in Multiple Myeloma patients?

**PERTINENT FINDINGS:** Thirty patients were prospectively included and had 18-FDG PET/CT and Whole-Body diffusion-weighted MRI, the latter being the gold standard for diffuse disease assessment. Machine learning analysis of Radiomics data, extracted from bone marrow PET/CT images, was able to reach high accuracy for the diagnosis of diffuse disease, superior to visual analysis.

**IMPLICATIONS FOR PATIENT CARE:** The conceived model could improve 18-FDG PET/CT overall diagnostic value for multiple myeloma disease assessment at baseline and therefore for subsequent evaluations of response to therapy.

## References

1. Regelin JC, Minnema MC, Terpos E, et al. Comparison of modern and conventional imaging techniques in establishing multiple myeloma-related bone disease: a systematic review. *Br J Haematol*. 2013;162:50-61.
2. Usmani SZ, Mitchell A, Waheed S, et al. Prognostic implications of serial 18-fluorodeoxyglucose emission tomography in multiple myeloma treated with total therapy 3. *Blood*. 2013;121:1819-1823.
3. Moreau P, Attal M, Caillot D, et al. Prospective Evaluation of Magnetic Resonance Imaging and [18F]Fluorodeoxyglucose Positron Emission Tomography-Computed Tomography at Diagnosis and Before Maintenance Therapy in Symptomatic Patients With Multiple Myeloma Included in the IFM/DFCI 2009 Trial: Results of the IMAJEM Study. *J Clin Oncol*. 2017;35:2911-2918.
4. Mouloupoulos LA, Gika D, Anagnostopoulos A, et al. Prognostic significance of magnetic resonance imaging of bone marrow in previously untreated patients with multiple myeloma. *Ann Oncol*. 2005;16:1824-1828.
5. Mouloupoulos LA, Dimopoulos MA, Kastiris E, et al. Diffuse pattern of bone marrow involvement on magnetic resonance imaging is associated with high risk cytogenetics and poor outcome in newly diagnosed, symptomatic patients with multiple myeloma: a single center experience on 228 patients. *Am J Hematol*. 2012;87:861-864.
6. Zamagni E, Nanni C, Patriarca F, et al. A prospective comparison of 18F-fluorodeoxyglucose positron emission tomography-computed tomography, magnetic resonance imaging and whole-body planar radiographs in the assessment of bone disease in newly diagnosed multiple myeloma. *Haematologica*. 2007;92:50-55.
7. Hillengass J, Usmani S, Rajkumar SV, et al. International myeloma working group consensus recommendations on imaging in monoclonal plasma cell disorders. *Lancet Oncol*. 2019;20:e302-e312.
8. Zamagni E, Nanni C, Mancuso K, et al. PET/CT Improves the Definition of Complete Response and Allows to Detect Otherwise Unidentifiable Skeletal Progression in Multiple Myeloma. *Clin Cancer Res*. 2015;21:4384-4390.
9. Mesguich C, Fardanesh R, Tanenbaum L, Chari A, Jagannath S, Kostakoglu L. State of the art imaging of multiple myeloma: comparative review of FDG PET/CT imaging in various clinical settings. *Eur J Radiol*. 2014;83:2203-2223.
10. Hatt M, Majdoub M, Vallieres M, et al. 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med*. 2015;56:38-44.

11. Orlhac F, Theze B, Soussan M, Boisgard R, Buvat I. Multiscale Texture Analysis: From 18F-FDG PET Images to Histologic Images. *J Nucl Med*. 2016;57:1823-1828.
12. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
13. Rajkumar SV, Dimopoulos MA, Palumbo A, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol*. 2014;15:e538-548.
14. Duvauferrier R, Valence M, Patrat-Delon S, et al. Current role of CT and whole body MRI in multiple myeloma. *Diagn Interv Imaging*. 2013;94:169-183.
15. Denis de Senneville B, Zachiu C, Ries M, Moonen C. EVolution: an edge-based variational method for non-rigid multi-modal image registration. *Phys Med Biol*. 2016;61:7377-7396.
16. Lafitte L, Zachiu C, Kerkmeijer LGW, Ries M, Denis de Senneville B. Accelerating multi-modal image registration using a supervoxel-based variational framework. *Phys Med Biol*. 2018;63:235009.
17. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res*. 2017;77:e104-e107.
18. Zwanenburg A, Leger S, Vallières M, Löck S, Initiative f. Image biomarker standardisation initiative - feature definitions. 2016.
19. Orlhac F, Soussan M, Chouahnia K, Martinod E, Buvat I. 18F-FDG PET-Derived Textural Indices Reflect Tissue-Specific Uptake Pattern in Non-Small Cell Lung Cancer. *PLoS One*. 2015;10:e0145063.
20. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-2830.
21. Breiman L. Random Forests. *Machine Learning*. 2001;45:5-32.
22. Nanni C, Versari A, Chauvie S, et al. Interpretation criteria for FDG PET/CT in multiple myeloma (IMPeTUs): final results. IMPeTUs (Italian myeloma criteria for PET USE). *Eur J Nucl Med Mol Imaging*. 2018;45:712-719.
23. Reuze S, Schernberg A, Orlhac F, et al. Radiomics in Nuclear Medicine Applied to Radiation Therapy: Methods, Pitfalls, and Challenges. *Int J Radiat Oncol Biol Phys*. 2018;102:1117-1142.
24. Aide N, Talbot M, Fruchart C, Damaj G, Lasnon C. Diagnostic and prognostic value of baseline FDG PET/CT skeletal textural features in diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2018;45:699-711.



- 25.** Li H, Xu C, Xin B, et al. (18)F-FDG PET/CT Radiomic Analysis with Machine Learning for Identifying Bone Marrow Involvement in the Patients with Suspected Relapsed Acute Leukemia. *Theranostics*. 2019;9:4730-4739.
- 26.** Hillengass J, Fechtner K, Weber MA, et al. Prognostic significance of focal lesions in whole-body magnetic resonance imaging in patients with asymptomatic multiple myeloma. *J Clin Oncol*. 2010;28:1606-1610.
- 27.** Rasche L, Alapat D, Kumar M, et al. Combination of flow cytometry and functional imaging for monitoring of residual disease in myeloma. *Leukemia*. 2019;33:1713-1722.

## Figures

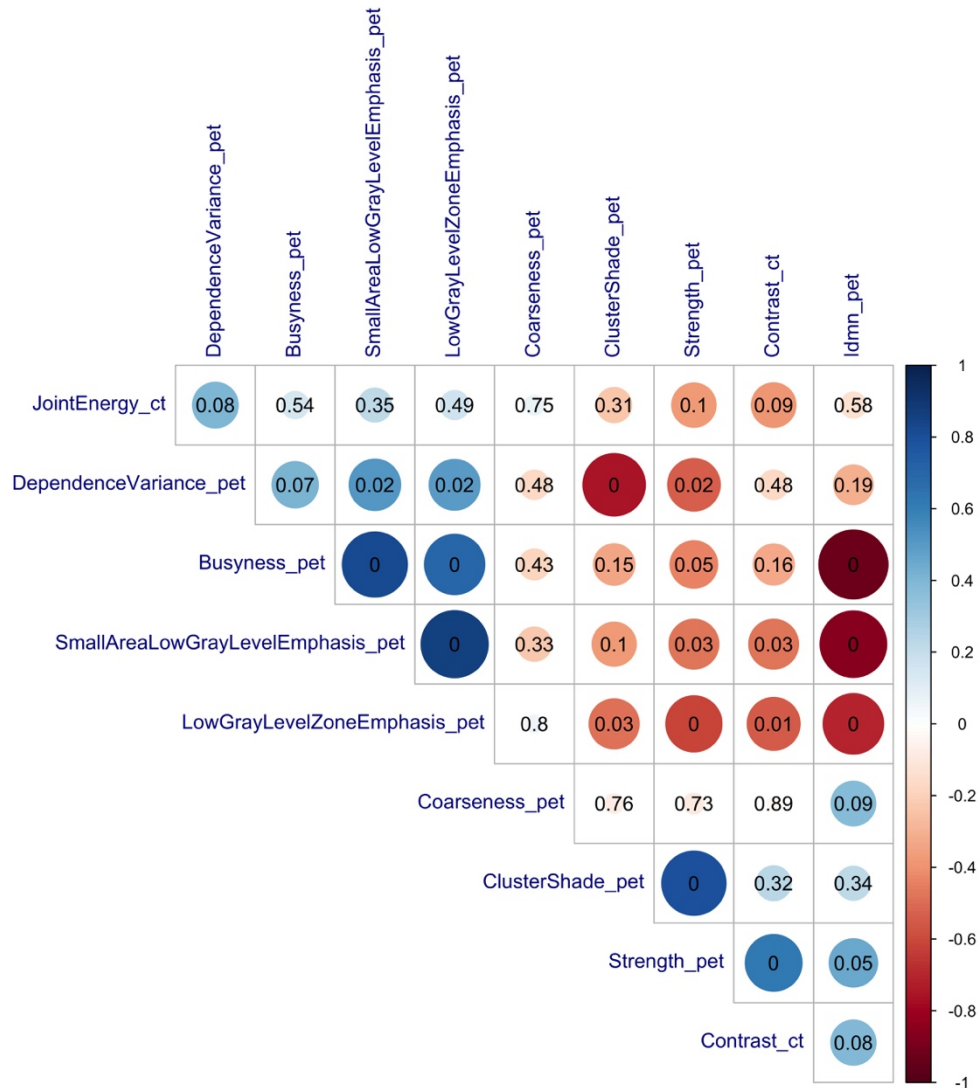


Figure 1. Correlation Matrix of the features pre-selected by random forest tree importance.

Circle diameter and color represent the magnitude of feature correlation. The  $P$  value of correlation test is showed at the center of each circle.

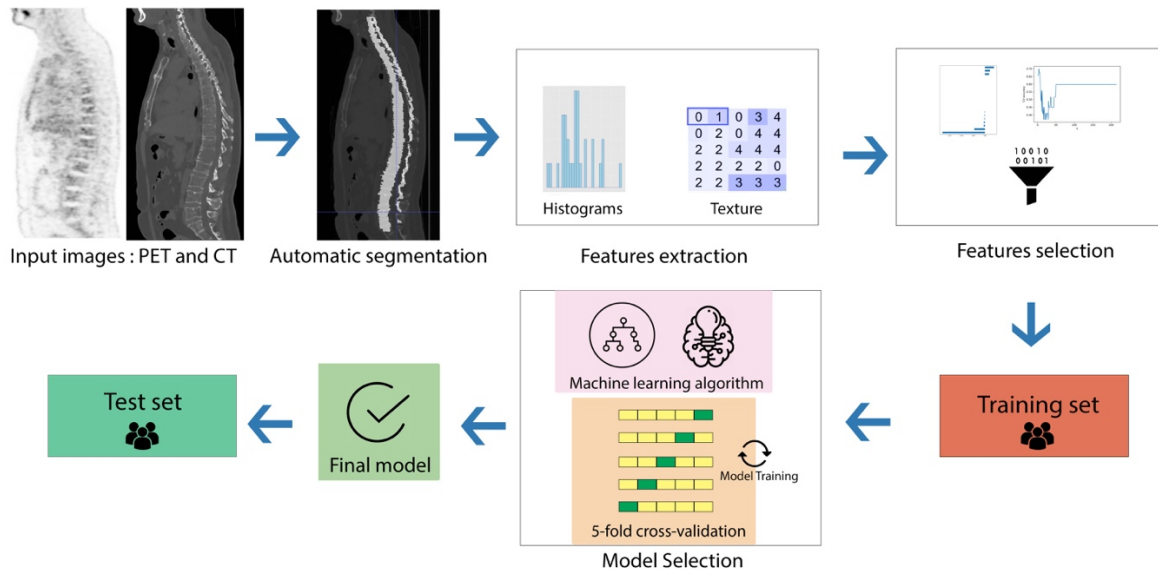


Figure 2. Flowchart of Radiomics analysis

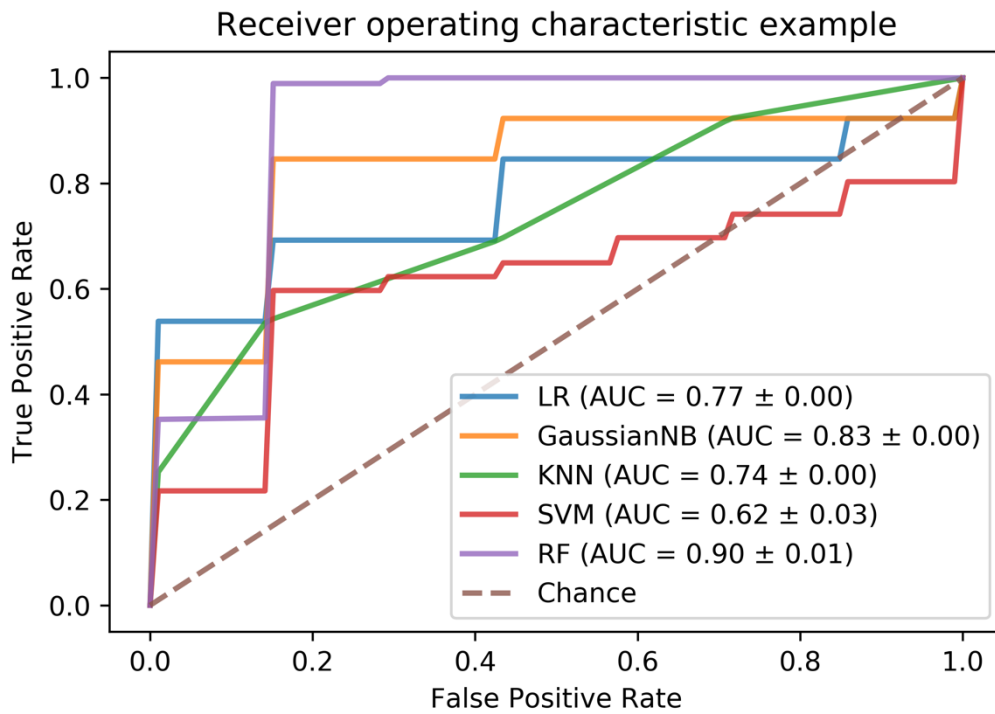


Figure 3. ROC curves of the different classifier's performance on the training set

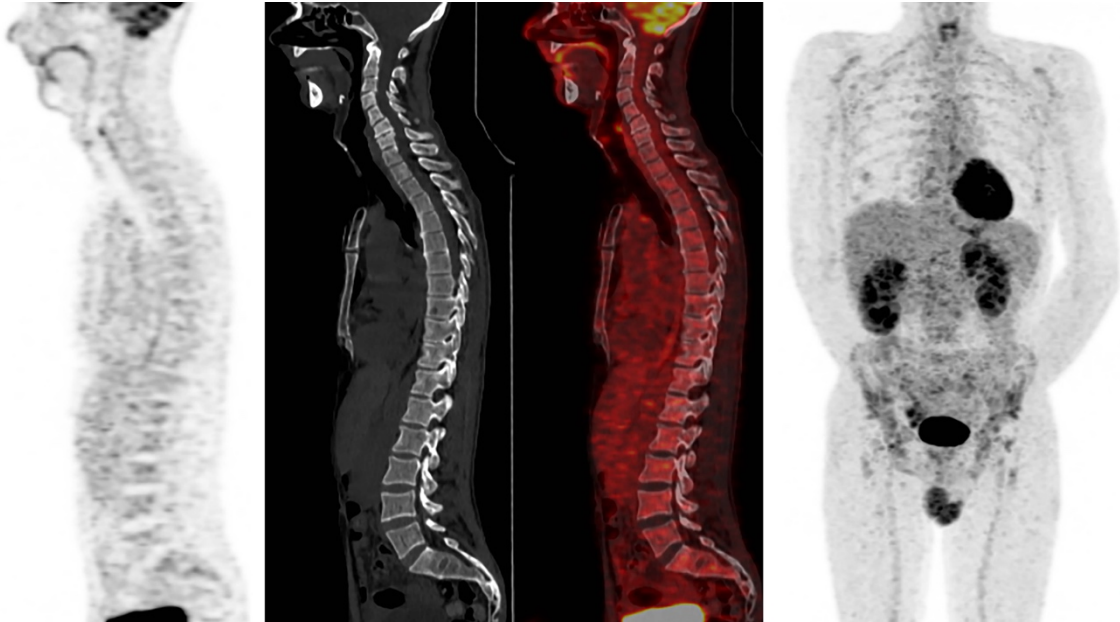


Figure 4. 54-year-old male with IgD lambda ISS1 Multiple Myeloma. Sagittal 18-FDG PET images show a mild diffuse uptake of the bone marrow of the spine without underlying bone lesions on the companion CT. Bone marrow biopsy showed 24% of clonal plasma cells. Whole-Body Diffusion-Weighted MRI was positive for diffuse disease with the presence of a severe diffuse infiltration. Visual analysis failed to detect bone marrow diffuse disease. The Machine Learning algorithm based on Radiomics data extracted from PET/CT images correctly predicted the presence of diffuse disease.

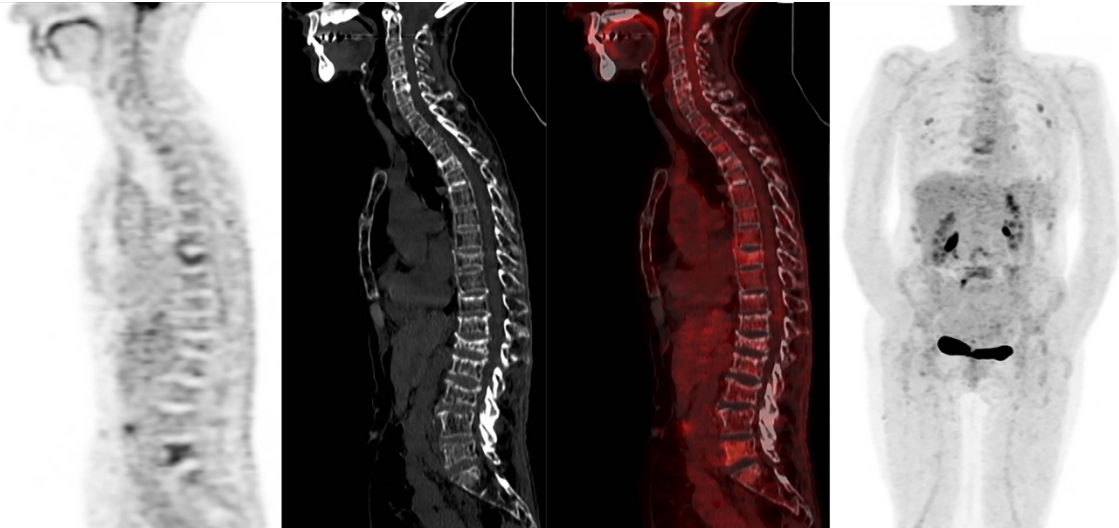


Figure 5. 62-year-old female with ISS1 kappa IgG Multiple Myeloma. Sagittal 18-FDG PET images show a mild and diffuse bone marrow uptake of the spine assorted with focal uptakes of few vertebral bodies. Sagittal CT images show multiple vertebra fractures and a diffuse alteration of bone density. Other focal uptakes are seen within the ribs. Bone marrow biopsy showed 41% of clonal plasma cells. Whole-Body Diffusion-Weighted MRI was negative for diffuse disease. Visual analysis of 18-FDG PET/CT images was wrongly positive. The machine learning algorithm correctly predicted the absence of diffuse disease.

## Tables

Table 1. Patient characteristics

Characteristic	Value <i>n</i> =30	Range or %
Median age	56	34–66
Sex		
Male	17	56
Type of myeloma		
IgG	12	40
IgA	5	17
IgD	1	3
Light chain	12	40
Anemia	19	63
Thrombocytopenia	3	10
Altered renal function	4	13
Hypercalcemia	3	10
Elevated LDH	13	43
Beta-2 microglobulin		
>3.5 mg/L	15	50
>5.5 mg/L	5	17
Median serum free light chain (mg/L)	343	3.2–17,600
R-ISS score		
1	15	50
2	9	30
3	6	20
Unfavorable cytogenetics	4	13
Median bone marrow plasma cell infiltration (%)	28	0–87

LDH, lactate dehydrogenase; R-ISS, Revised-International Staging System.

Table 2. Visual reader performance (on the 20 patients from the training set)

	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	Accuracy (95%CI)
<b>Reader 1</b>	0.62 (0.35-0.88)	0.43 (0.06-0.80)	0.67 (0.40-0.93)	0.38 (0.04-0.71)	0.55 (0.32-0.77)
<b>Reader 2</b>	0.62 (0.35-0.88)	0.71 (0.38-1.00)	0.80 (0.55-1.00)	0.50 (0.19-0.81)	0.65 (0.41-0.85)
<b>Consensus reading</b>	0.67 (0.45-0.88)	0.75 (0.51-0.99)	0.80 (0.60-1.00)	0.60 (0.35-0.85)	0.70 (0.50-0.85)

Table 3. Classifier diagnostic performances on the training set and on the independent test set with the 2 selected radiomics features.

<b>Classifier</b>	<b>Sensitivity (95% CI)</b>	<b>Specificity (95% CI)</b>	<b>PPV (95% CI)</b>	<b>NPV (95% CI)</b>	<b>F1-Score (95% CI)</b>	<b>Accuracy (95% CI)</b>
<b>Training Set*</b>						
Logistic Regression	0.82 (0.81-0.83)	0.76 (0.74-0.77)	0.89 (0.87-0.92)	0.65 (0.58-0.72)	0.83 (0.81-0.85)	0.80 (0.79-0.81)
Gaussian Naive-Bayes	0.56 (0.53-0.60)	0.85 (0.82-0.88)	0.80 (0.75-0.85)	0.54 (0.45-0.62)	0.63 (0.60-0.66)	0.66 (0.63-0.69)
K-Nearest Neighbors	0.79 (0.78-0.80)	0.82 (0.78-0.86)	0.91 (0.88-0.94)	0.69 (0.62-0.76)	0.82 (0.81-0.84)	0.80 (0.78-0.82)
Support Vector Machine	0.78 (0.77-0.81)	0.77 (0.73-0.81)	0.88 (0.85-0.91)	0.65 (0.58-0.71)	0.81 (0.78-0.83)	0.78 (0.76-0.80)
Random Forest	0.93 (0.91-0.96)	0.86 (0.84-0.88)	0.94 (0.93-0.96)	0.85 (0.76-0.94)	0.93 (0.92-0.94)	0.91 (0.90-0.92)
<b>Test Set</b>						
Logistic Regression	0.80	0.40	0.57	0.67	0.67	0.60
Gaussian Naive-Bayes	0.80	0.80	0.80	0.80	0.80	0.80
K-Nearest Neighbors	0.67	0.57	0.40	0.80	0.67	0.60
Support Vector Machine	0.50	0.50	0.40	0.60	0.55	0.50
Random Forest	1.00	0.71	0.60	1.00	0.75	0.80

PPV, Positive Predictive Value; NPV, Negative Predictive Value.

\*Results of the training set were obtained after a 5-fold shuffled cross-validation iterated 100 times with a different initialization seed.