

Radiomics analysis of [¹⁸F] fluorodeoxyglucose-avid thyroid incidentalomas improves risk stratification and selection for clinical assessment

Luca Ceriani, Lisa Milan, Camilla Virili, Luciano Cascione, Gaetano Paone, Pierpaolo Trimboli, Luca Giovanella.

Luca Ceriani, MD. Clinic for Nuclear Medicine and Molecular Imaging. Imaging Institute of Southern Switzerland. Via Tesserete 46, 6900 Lugano, Switzerland and Institute of Oncology Research, Faculty of Biomedical Sciences, Università della Svizzera italiana. Via Vela 6, 6500 Bellinzona, Switzerland. E-mail: Luca.Ceriani@eoc.ch

Lisa Milan, MPE. Clinic for Nuclear Medicine and Molecular Imaging. Imaging Institute of Southern Switzerland. Via Tesserete 46, 6900 Lugano, Switzerland. E-mail: Lisa.Milan@eoc.ch

Camilla Virili, MD. Department of Medico-Surgical Sciences and Biotechnologies, Sapienza University of Rome, Corso della Repubblica 79, 04100 Latina, Italy. E-mail: camilla.virili@uniroma1.it

Luciano Cascione, PhD. Institute of Oncology Research, Faculty of Biomedical Sciences, Università della Svizzera italiana, Via Vela 6, 6500 Bellinzona, Switzerland, and SIB Swiss Institute of Bioinformatics, Quartier Sorge - Batiment Amphipole, 1015, Lausanne, Switzerland. E-mail: luciano.cascione@ior.usi.ch

Gaetano Paone, MD. Clinic for Nuclear Medicine and Molecular Imaging. Imaging Institute of Southern Switzerland. Via Tesserete 46, 6900 Lugano, Switzerland. E-mail: Gaetano.Paone@eoc.ch

Pierpaolo Trimboli, Prof. Clinic for Nuclear Medicine and Molecular Imaging. Imaging Institute of Southern Switzerland. Via Tesserete 46, 6900 Lugano, Switzerland and Competence Centre for Thyroid Diseases, Ente Ospedaliero Cantonale, Via Ospedale 12, 6500 Bellinzona, Switzerland and Faculty of Biomedical Sciences, Università della Svizzera Italiana (USI), Via Buffi 13, 6900 Lugano, Switzerland. E-mail: Pierpaolo.Trimboli@eoc.ch

Luca Giovanella, Prof. Clinic for Nuclear Medicine and Molecular Imaging. Imaging Institute of Southern Switzerland. Via Tesserete 46, 6900 Lugano, Switzerland and Competence Centre for Thyroid Diseases, Ente Ospedaliero Cantonale, Via Ospedale 12, 6500

Bellinzona, Switzerland and Clinic for Nuclear Medicine, University Hospital and University of Zurich, Rämistrasse 100, 8091 Zurich, Switzerland. E-mail: luca.giovanella@eoc.ch

Running title: Radiomics of FDG-avid thyroid incidentalomas

Key words: [^{18}F]FDG-PET-CT; Thyroid incidentaloma; radiomics; PET metrics; predictive model.

Abstract

Background: [¹⁸F]-Fluorodeoxyglucose (FDG)-avid thyroid lesions incidentally detected on positron emission tomography/computed tomography (PET/CT) scans represent a tumor lesion in about 30% of cases. The present study evaluated the ability of PET metrics and radiomics features to predict final diagnosis of [¹⁸F] FDG thyroid incidentalomas (TIs).

Methods: A total of 104 patients with 107 TIs were retrospectively studied; 30 nodules (28%) were diagnosed as malignant. After volumetric segmentation of each thyroid lesion, metabolic tumor volume, total lesion glycolysis (TLG), standardized uptake values (SUVs), and metabolic heterogeneity were estimated, and 107 radiomics features were extracted following a standard protocol.

Results: Metabolic tumor volumes, TLG, SUVmax, SUVmean, and SUVpeak among functional PET parameters, and GLCM_InverseDifferenceMoment, shape_Sphericity, GLCM_SumSquares, firstorder_Maximum2DDiameterSlice, firstorder_Energy, and GLCM_Contrast among nonredundant radiomics features, showed significantly different values between malignant and benign TIs (Mann Whitney-U, $P < 0.01$ for all). Univariate logistic regression revealed these parameters demonstrated good ability to predict final diagnosis of TIs ($P < 0.02$ for all). Shape_Sphericity was the best predictor classifying 82% of TIs correctly ($P < 0.0001$). Only TLG, SUVmax, and shape_Sphericity retained significance ($P < 0.0001$) by multivariate analysis. Malignant lesion prevalence increased from 7% to 100% in accordance with the number (score, 0–3) of the three positive parameters present (χ^2 trend, $P < 0.0001$). A score of 0 excludes malignant TIs with a negative predictive value of 93%, while a score of 3 predicted malignancy with a positive predictive value of 100%.

Conclusions: PET metrics and radiomics analysis can improve identification of [¹⁸F]FDG-avid TIs at high risk of malignancy. A model based on TLG, SUVmax, and shape_Sphericity may allow prediction of a final diagnosis, providing useful information for the management of TIs.

Introduction

Incidentally discovered [^{18}F]-fluorodeoxyglucose (FDG)-avid thyroid incidentalomas (TIs) are reported in about 1.5%–2% of all positron emission tomography/computed tomography (PET/CT) scans performed for nonthyroid diseases (1-3), and about 30% are malignant (1, 2, 4, 5). Unfortunately, neither visual analysis nor PET-derived parameters, such as standardized uptake values (SUVs), metabolic tumor volume (MTV), and total lesion glycolysis (TLG), reliably discriminate malignant from benign [^{18}F]FDG-avid TIs (3, 6-11). Ultrasound (US) assessment is pivotal in detection and exclusion of thyroid cancers among patients with thyroid nodules, and recently, several US risk stratification systems have been used in clinical practice (12). However, studies focusing only on the performance of US have provided sparse results (13-15), and limited preliminary data are currently available on US risk stratification systems used in patients with [^{18}F]FDG-TIs (16, 17).

International clinical guidelines strongly recommend use of fine-needle aspiration cytology to ascertain [^{18}F]FDG-avid TIs greater than 1 cm in greatest diameter (5). However, the prognosis of oncology patients is dictated by the underlying malignancy rather than by incidentally detected thyroid cancers in most cases. Accordingly, the costs, risks, and anxiety associated with incidentally detected [^{18}F]FDG-avid thyroid carcinomas should be carefully balanced by primary malignancy (18-20).

Accurate PET-derived criteria would be helpful to directly stratify the malignancy risk of [^{18}F]FDG-TIs during PET/CT interpretation and reduce the need for further evaluation and the associated risks, costs, and patient discomfort. Recent preliminary data have suggested that heterogeneous [^{18}F]FDG uptake within thyroid nodules (i.e., “metabolic heterogeneity”) and PET/CT texture analysis are promising approaches to stratify the risk of a malignant nodule in patients with [^{18}F]FDG-avid incidentalomas, refine selection of patients for cytology referral, and reduce the need for further investigations (21-23). The present study investigated the role of PET-derived parameters and radiomics features to characterize [^{18}F]FDG-TIs in a large series of patients with cancer and generated a predictive model for discriminating malignant from benign [^{18}F]FDG-TIs based on PET/CT-derived data.

Materials and methods

Study design and patient selection

Initially, our facility's database was searched for all patients who underwent [¹⁸F]FDG-PET/CT due to non-thyroid oncological indications from October 2013 to November 2018. Then, all cases with abnormal thyroid [¹⁸F]FDG uptake were retrieved. After exclusion of cases with diffuse [¹⁸F]FDG thyroid uptake, patients with focal [¹⁸F]FDG uptake were selected. The study inclusion criteria were: 1) a histological diagnosis, 2) benign cytological report in the absence of postoperative histology, and 3) scintigraphically-confirmed autonomously functioning thyroid nodules that received curative treatment with different modalities (24). All fine-needle aspiration cytology was reported according to the Italian Consensus for the Classification and Reporting of Thyroid Cytology [i.e., TIR1 (inadequate), TIR2 (benign), TIR3 (indeterminate), TIR4 (suspicious for malignancy), and TIR5 (malignancy)] (25).

[¹⁸F]FDG-PET/CT image acquisition

[¹⁸F]FDG-PET/CT whole-body images were acquired following a standard protocol in accordance with European Association of Nuclear Medicine guidelines (26) using two integrated PET/CT scanners of the same model (both Siemens Biograph mCT 40, Siemens, Erlangen, Germany). Scanning was performed 60 ± 5 min after intravenous injection of 3 MBq/kg of [¹⁸F]FDG. All patients fasted for at least 6 hours prior to image acquisition and their fasting blood glucose was <160 mg/dL. CT scans obtained with a low-dose protocol were used for attenuation-correction of PET images. Images were reconstructed with 3-dimensional ordered subset expectation maximization iterative algorithms (3 iterations and 21 subsets), including Time-of-Flight and Point-Spread Function corrections.

[¹⁸F]FDG -PET/CT metrics

Reconstructed [¹⁸F]FDG-PET/CT images were analyzed with dedicated software (MM Oncology, Syngo.via, Siemens, Erlangen, Germany). An expert nuclear medicine physician (LC) contoured all thyroid focal PET findings with a 3-dimensional fixed threshold algorithm. The MTV of each TI was automatically estimated setting the mean SUV of the contralateral lobe as threshold to identify the borders between pathological and normal tissue. Then, maximum, mean, and peak SUVs (SUV_{max}, SUV_{mean}, and SUV_{peak}, respectively) of the lesion were measured automatically. The TLG was calculated as the

product of the SUVmean and MTV (27). Metabolic heterogeneity was assessed using the cumulative SUV volume histograms (CSH) method (28), as applied in previous studies (29-34). The area under the curve of the CSH is a quantitative index of tracer uptake heterogeneity where lower values correspond to increased heterogeneity (35). Metabolic heterogeneity was also evaluated by estimating the coefficient of variation of intra-lesion [^{18}F]FDG uptake, which was calculated as the standard deviation of the SUVs divided by the SUVmean of the segmented lesion (30, 36).

Radiomics features

Radiomics features were extracted from each segmented volume using PyRadiomics software package version 2.2.0 (37). In order to standardize the process of extraction, gray level intensities and voxel dimensions of the original images were preliminary resampled following the Image Biomarkers Standardization Initiative recommendations (38-41). From the segmented volumes, 107 standardized features evaluating different metabolic characteristics of the lesion were initially extracted. They included 14 shape-based features, 18 first-order statistics features and 75 matrix-based features (24 gray-level co-occurrence matrix (GLCM), 16 gray-level run length matrix (GLRLM), 16 gray-level size zone matrix (GLSZM), 5 neighboring gray tone difference matrix (NGTDM), and 14 gray-level dependence matrix (GLDM) based). Meaning and mathematical description of these radiomics features are reported in detail in the PyRadiomics documentation [<https://pyradiomics.readthedocs.io/en/latest/features.html>].

The shape based features measure the dimensions and depict the geometric properties of the lesion. The first order statistics features describe the presence of areas with statistically different metabolic activity without taking into account their distribution within the lesion, while the matrix-based features analyze spatial distribution of these areas and their mutual relationship. In the last group, the analysis is performed applying multiple mathematical models to the data organized in the form of matrices.

Since the image were acquired with two different scanners, the first analytic step was aimed at evaluating the inter-scanner reproducibility of the radiomics features. For each feature, the distributions of the values extracted by the two sets of images were compared. The features with different distributions were excluded from the final analysis

to avoid a potential bias, while those with statistically proved inter-scanner reproducibility were tested as potential predictors of the final diagnosis of TIs.

Reference standard

Malignant lesions described throughout the manuscript were thyroid cancers confirmed by surgical pathology examination. Benign lesions were defined as follows: histologically-confirmed benign nodules (42), cytologically benign nodules (25) with clinical and US follow-up ≥ 3 years, and scintigraphically-confirmed autonomously functioning thyroid nodules (43).

Ethics

This retrospective study on existing patient data and images was reviewed and approved by our Advisory Research Board and the Ethic Committee of Canton Tessin, Switzerland (ref. BASEC 2018-00685, CE TI 3351). The requirement for informed consent was waived. All procedures were in accordance with the ethical standards of the 1964 Helsinki Declaration and its later amendments.

Statistical analyses

Quantitative variables were expressed as medians and interquartile ranges. Medians were compared by Mann Whitney-U test. Differences between categorical data frequency were assessed by χ^2 test. Reproducible radiomics features were identified by comparing results from the two PET-CT scanners, applying the Mann-Whitney U-test with false discovery rate correction. Nonredundant and uncorrelated radiomics features were selected using the correlation-based feature selection algorithm, a supervised method for reducing the dimensionality. Continuous parameters were analyzed as dichotomized variables using receiver-operating characteristic curves and Youden's coefficient method to estimate the optimal cut-off point for discriminating malignant from benign TIs. Univariate logistic regression analysis was applied to assess the relationship between each parameter and the final diagnosis. Logistic stepwise regression function was used for multivariate analysis. In order to evaluate the robustness and the generalization of our model and considering the sample size of the study a 1000-resampled bootstrapping was performed as cross validation procedure. A $P < 0.05$ was considered statistically significant. Negative and positive predictive values were calculated. Statistical analyses were conducted using RStudio statistical software package version 1.2.503 (RStudio Inc., Boston, MA, USA;

<http://www.rstudio.com/>) and MedCalc Statistical Software version 15.8 (MedCalc Software bvba, Ostend, Belgium).

Results

Characteristics of patients

Among the 12,652 consecutive patients who underwent [¹⁸F]FDG-PET/CT during the study period, 333 (2.6%) TIs were recorded. After exclusion of cases with diffuse [¹⁸F]FDG thyroid uptake patterns ($n = 146$), 187 (57%) [¹⁸F]FDG-TIs were included. According to the selection criteria, 107 focal TIs from 104 patients (67 females and 37 males; median age, 65 years) were selected for the present analysis (Figure 1). The median TSH was 1.5 mIU/L (0.96–2.1 mIU/L). Lesions had a median size of 17 mm (12–26 mm). As summarized in Table 1, 30/107 (28%) TIs were malignant (26 primary thyroid carcinomas, 1 non-Hodgkin lymphoma, and 3 metastatic lesions from sarcoma, renal and esophageal cancer, respectively) and 77 (72%) were benign (14 autoimmune thyroiditis, 1 autonomously functioning thyroid nodules, and 62 hyperplastic nodules).

PET metrics and radiomics analysis

Among functional and volumetric PET-derived parameters, MTV, TLG, SUVmax, SUVmean, and SUVpeak values were significantly increased in malignant compared to benign TIs (Table 2). After dichotomization, such parameters reliably predicted the final diagnosis on univariate analysis. TLG was the best predictor, correctly classifying 79% of lesions (univariate logistic regression, $P < 0.0001$). Specifically, low TLG values were associated with benign lesions and had a negative predictive value of 84%, while higher TLG values had a positive predictive value of 65% for malignant nodules. On the other hand, metabolic heterogeneity assessed either by area under the curve-CSH or coefficient of variation was not associated with the final diagnosis due to the wide overlap of values between the two groups.

Among the 107 extracted radiomics features, only 54 were statistically reproducible between the two PET/CT scanners. Among these, the correlation-based feature selection algorithm selected 6 nonredundant and uncorrelated radiomics features as potential predictors of the final diagnosis. The 6 selected features are summarized in Table 2. The shape_Sphericity and the shape_Maximum2DDiameterSlice are two morphological

features measuring the degree of spherical shape and the largest dimension in the axial plane of the lesion, respectively.

The firstorder_Energy is a gray levels histogram based feature measuring the intensity of voxel values (i.e. the metabolic activity) in the image. The GLCM_Contrast, GLCM_inverse difference moment and GLCM_SumSquares are three textural features belonging to the group of the GLCM based features that describe with different tools the degree of metabolic heterogeneity of the TI.

Their ability to distinguish malignant from benign TIs was significant by Mann-Whitney U-test (Table 2) and univariate logistic regression of dichotomized data (Table 3). Shape_Sphericity had the highest accuracy, correctly classifying 82% of TIs (univariate logistic regression, $P < 0.0001$). TIs closer to the spherical shape (i.e. with higher values of sphericity) had a lower risk of malignancy with negative predictive values of 82%, while the lack of sphericity identified a tumor lesion with positive predictive value of 82% (Table 3). In multivariate stepwise logistic regression analysis including the previously selected 5 functional PET parameters and 6 radiomics features, TLG, SUVmax, and shape_Sphericity retained statistical significance ($P < 0.0001$; Table 4). The 1000-resampled bootstrapping method validated the multivariate analysis results with an optimism-corrected area under the curve of 0.830 ($P < 0.0001$).

Multiparametric predictive model

Based on these results, we tested a predictive model integrating the dichotomized TLG, SUVmax, and shape_Sphericity. In this model, four subgroups of lesions were defined by a score from 0 to 3 reproducing the number of positive imaging-derived parameters. This model was accurate for stratifying the risk of TI malignancy (χ^2 test, $P < 0.0001$). The prevalence of malignant TIs increased from 7% (3/41) for lesions with a score of 0 to 21% (9/42), 60% (9/15), and 100% (9/9) for lesions with scores of 1, 2, and 3, respectively (χ^2 trend, $P < 0.0001$). Risk of malignancy only remained comparable to that of the overall population in the group with a score of 1 (χ^2 test $P = 0.503$), while the group with a score of 0 was significantly lower (χ^2 test, $P = 0.0114$) and groups with scores of 2 and 3 were significantly higher (χ^2 test, $P = 0.0282$ and < 0.0001 , respectively). In summary, a score of

0 identified benign TIs with a negative predictive value of 93%, while scores of 2 and 3 predicted malignancy with positive predictive values of 60% and 100%, respectively.

Discussion

Assessment of [^{18}F]FDG TIs is a critical challenge for oncology patients as the reported thyroid cancer prevalence ranges from 20% to 40%. Furthermore, patients undergoing [^{18}F]FDG-PET/CT are generally affected by aggressive nonthyroid cancers that commonly dictates their prognosis. Thus, developing a PET/CT-based strategy to stratify the risk of malignancy in FDG-TIs may assist nuclear medicine physicians during PET/CT reporting and provide patient-specific risk-stratification. In turn, additional referrals and investigations could be reduced, positively impacting patient quality of life and reducing overall costs.

Previous studies have reported increased SUVmax and SUVmean values and higher PET volume-based measurements (MTV and TLG) in malignant TIs, suggesting a potential role for these parameters, alone or in combination, as predictors of TI malignancy (7, 9, 10, 22). Others, however, have reported no significant association of the same parameters with TI malignancy (1, 6, 8, 11). More recently, a different approach based on assessment of [^{18}F]FDG distribution within thyroid nodules has been proposed (21-23), and preliminary data suggest that high metabolic heterogeneity may identify malignant TIs better than conventional PET metrics (21). Sollini *et al.* reported interesting preliminary results by evaluating histogram-based and matrix-based features by textural analysis (22, 23).

To the best of our knowledge, our study is the first attempt to integrate textural features and conventional PET metrics in a multiparametric predictive model for TIs. The present data confirmed the predictive power of conventional PET metrics parameters and identified 6 independent radiomics features describing shape, heterogeneity, and intensity of lesion tracer uptake as additional tools for discriminating benign from malignant TIs. As the main result, a predictive model combining the three independent PET-derived predictive parameters (i.e., TLG, SUVmax, and shape_Sphericity) proved to be highly accurate for stratifying the risk of malignancy of [^{18}F]FDG-avid TIs. In particular, all triple-positive TIs were malignant (positive predictive value, 100%), while 93% of triple-negative cases were benign.

The present retrospective, single-center study has some limitations. First, in some cases, the severity of the primary malignancy could have prevented a complete diagnostic work-

up of the TI and the exclusion of these patient in our analysis. However, the prevalence of thyroid cancer among the study cohort was in line with current rates reported in the literature. Second, a validated threshold value to segment thyroid PET-positive thyroid nodules has not been defined yet. However, our approach was based on arbitrarily selecting the SUVmean of the contralateral lobe to define the actual volume of TIs independently from their metabolic activity increasing accuracy and reproducibility of radiomics analysis. Third, we analyzed only lesions larger than 1.1 cm, consistent with the current recommendation to not perform fine-needle aspiration cytology in nodules, included FDG-active ones, less than 1 cm in largest diameters (5). Thus, our findings may not be applicable to smaller TIs. Fourth, even if the present series is larger than those previously reported, validation of the current model in a prospective study including a larger number of cases is warranted to confirm these results.

Finally, we did not compare PET data with ultrasound pattern in our series as ultrasound examinations were performed by different sonographers in different centers precluding any reliable comparison. On the other hand, it should be noted that, even if the potential role of ultrasound in this setting was reported (17) current clinical guidelines still recommend fine-needle aspiration cytology in all FDG-active thyroid incidentalomas > 1 cm (5). Accordingly, the aim of this study was to test the possibility of better characterizing metabolic PET information to guide the need for fine needle aspiration in patients with cancer independent of other evaluation. Our PET/CT-based model was able to predict the risk of malignancy of PET TIs. Such approach may potentially avoid unnecessary additional procedures. The potential improvement generated by the integration between PET and ultrasound risk factors remains of course of high interest and deserves to be explored in future studies.

In conclusion, the present multiparametric PET/CT-based radiomics model showed good performance in stratifying the risk of malignancy of [¹⁸F]FDG-avid TIs and may be useful for reducing the number of fine-needle aspiration cytology referrals in patients with cancer.

Author Disclosure Statements

The authors have no competing financial interests to declare.

Corresponding author:

Luca Ceriani, MD

Clinic for Nuclear Medicine and Molecular Imaging

Imaging Institute of Southern Switzerland

Via Tesserete 46, 6900 Lugano, Switzerland

E-mail: luca.ceriani@eoc.ch

Telephone: (+41) 91 811 6443. Fax: (+41) 91 811 6444.

References

1. Bertagna F, Treglia G, Piccardo A, Giubbini R 2012 Diagnostic and clinical significance of F-18-FDG-PET/CT thyroid incidentalomas. *J Clin Endocrinol Metab* **97**:3866-3875.
2. Treglia G, Giovanella L, Bertagna F, Di Franco D, Salvatori M 2013 Prevalence and risk of malignancy of thyroid incidentalomas detected by (18)f-fluorodeoxyglucose positron-emission tomography. *Thyroid* **23**:124-126.
3. Signore G, Albano D, Giovanella L, Bertagna F, Treglia G 2019 Evidence-based data about prevalence and risk of malignancy of thyroid incidentalomas detected by different PET radiopharmaceuticals. *Current radiopharmaceuticals*.
4. Treglia G, Bertagna F, Sadeghi R, Verburg FA, Ceriani L, Giovanella L 2013 Focal thyroid incidental uptake detected by (1)(8)F-fluorodeoxyglucose positron emission tomography. Meta-analysis on prevalence and malignancy risk. *Nuklearmedizin Nuclear medicine* **52**:130-136.
5. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, Schuff KG, Sherman SI, Sosa JA, Steward DL, Tuttle RM, Wartofsky L 2016 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* **26**:1-133.
6. Sager S, Vatankulu B, Sahin OE, Cınaral F, Uslu L, Baran A, Ozturk T, Sönmezoglu K 2018 Clinical significance of standardized uptake values in thyroid incidentaloma discovered by F-18 fluorodeoxyglucose positron emission tomography/computed tomography. *J Cancer Res Ther* **14**:989-993.
7. Shi H, Yuan Z, Yuan Z, Yang C, Zhang J, Shou Y, Zhang W, Ping Z, Gao X, Liu S 2018 Diagnostic Value of Volume-Based Fluorine-18-Fluorodeoxyglucose PET/CT Parameters for Characterizing Thyroid Incidentaloma. *Korean J Radiol* **19**:342-351.

8. Thuillier P, Bourhis D, Roudaut N, Crouzeix G, Alavi Z, Schick U, Robin P, Kerlan V, Salaun P-Y, Abgral R 2019 Diagnostic Value of FDG PET-CT Quantitative Parameters and Deauville-Like 5 Point-Scale in Predicting Malignancy of Focal Thyroid Incidentaloma. *Front Med (Lausanne)* **6**:24-24.
9. Kim BH, Kim S-J, Kim H, Jeon YK, Kim SS, Kim IJ, Kim YK 2013 Diagnostic value of metabolic tumor volume assessed by 18F-FDG PET/CT added to SUVmax for characterization of thyroid 18F-FDG incidentaloma. *Nucl Med Commun* **34**:868-876.
10. Soelberg KK, Bonnema SJ, Brix TH, Hegedüs L 2012 Risk of malignancy in thyroid incidentalomas detected by 18F-fluorodeoxyglucose positron emission tomography: a systematic review. *Thyroid* **22**:918-925.
11. Ho T-Y, Liou M-J, Lin K-J, Yen T-C 2011 Prevalence and significance of thyroid uptake detected by ¹⁸F-FDG PET. *Endocrine* **40**:297-302.
12. Castellana M, Castellana C, Treglia G, Giorgino F, Giovanella L, Russ G, Trimboli P 2019 Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. A meta-analysis. *J Clin Endocrinol Metab.*
13. Lin YH, Tsai YC, Lin KJ, Der Lin J, Wang CC, Chen ST 2019 Computer-Aided Diagnostic Technique in 2-Deoxy-2-[(18)F]fluoro-D-glucose-Positive Thyroid Nodule: Clinical Experience of 74 Non-thyroid Cancer Patients. *Ultrasound in medicine & biology* **45**:108-121.
14. Chung SR, Choi YJ, Suh CH, Kim HJ, Lee JJ, Kim WG, Sung TY, Lee YM, Song DE, Lee JH, Baek JH 2018 Thyroid Incidentalomas Detected on (18)F-Fluorodeoxyglucose Positron Emission Tomography with Computed Tomography: Malignant Risk Stratification and Management Plan. *Thyroid* **28**:762-768.
15. Beech P, Lavender I, Jong I, Soo G, Ramdave S, Chong A, Nandurkar D 2016 Ultrasound stratification of the FDG-avid thyroid nodule. *Clinical radiology* **71**:164-169.
16. Yoon JH, Cho A, Lee HS, Kim EK, Moon HJ, Kwak JY 2015 Thyroid incidentalomas detected on 18F-fluorodeoxyglucose-positron emission tomography/computed tomography: Thyroid Imaging Reporting and Data System (TIRADS) in the diagnosis and management of patients. *Surgery* **158**:1314-1322.

17. Trimboli P, Paone G, Treglia G, Virili C, Ruberto T, Ceriani L, Piccardo A, Giovanella L 2018 Fine-needle aspiration in all thyroid incidentalomas at (18) F-FDG PET/CT: Can EU-TIRADS revise the dogma? *Clinical endocrinology* **89**:642-648.
18. Poller DN, Bongiovanni M, Trimboli P 2020 Risk of malignancy in the various categories of the UK Royal College of Pathologists Thy terminology for thyroid FNA cytology: A systematic review and meta-analysis. *Cancer cytopathology* **128**:36-42.
19. Trimboli P, Fulciniti F, Paone G, Barizzi J, Piccardo A, Merlo E, Mazzucchelli L, Giovanella L 2020 Risk of Malignancy (ROM) of Thyroid FNA Diagnosed as Suspicious for Malignancy or Malignant: an Institutional Experience with Systematic Review and Meta-Analysis of Literature. *Endocrine pathology* **31**:52-56.
20. Pattison DA, Bozin M, Gorelik A, Hofman MS, Hicks RJ, Skandarajah A 2018 (18)F-FDG-Avid Thyroid Incidentalomas: The Importance of Contextual Interpretation. *J Nucl Med* **59**:749-755.
21. Kim S-J, Chang S 2015 Predictive value of intratumoral heterogeneity of F-18 FDG uptake for characterization of thyroid nodules according to Bethesda categories of fine needle aspiration biopsy results. *Endocrine* **50**:681-688.
22. Sollini M, Cozzi L, Pepe G, Antunovic L, Lania A, Di Tommaso L, Magnoni P, Erba PA, Kirienko M 2017 [(18)F]FDG-PET/CT texture analysis in thyroid incidentalomas: preliminary results. *Eur J Hybrid Imaging* **1**:3-3.
23. Sollini Martina, Kirienko Margarita, Magnoni Paola, Cozzi Luca , di Tommaso Luca, Lania Andrea Gerardo, Arturo C 2019 [18F]FDG-PET/CT texture analysis in thyroid "hot" nodules: results of the validation cohort. *Biomedical Journal of Scientific & Technical Research* **18**:13344-13350.
24. Giovanella L, Piccardo A, Pezzoli C, Bini F, Ricci R, Ruberto T, Trimboli P 2018 Comparison of high intensity focused ultrasound and radioiodine for treating toxic thyroid nodules. *Clinical endocrinology* **89**:219-225.
25. Nardi F, Basolo F, Crescenzi A, Fadda G, Frasoldati A, Orlandi F, Palombini L, Papini E, Zini M, Pontecorvi A, Vitti P 2014 Italian consensus for the classification and reporting of thyroid cytology. *Journal of endocrinological investigation* **37**:593-599.

26. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, Verzijlbergen FJ, Barrington SF, Pike LC, Weber WA, Stroobants S, Delbeke D, Donohoe KJ, Holbrook S, Graham MM, Testanera G, Hoekstra OS, Zijlstra J, Visser E, Hoekstra CJ, Pruim J, Willemsen A, Arends B, Kotzerke J, Bockisch A, Beyer T, Chiti A, Krause BJ, European Association of Nuclear M 2015 FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging* **42**:328-354.
27. Larson SM, Erdi Y, Akhurst T, Mazumdar M, Macapinlac HA, Finn RD, Casilla C, Fazzari M, Srivastava N, Yeung HW, Humm JL, Guillem J, Downey R, Karpeh M, Cohen AE, Ginsberg R 1999 Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. *The Visual Response Score and the Change in Total Lesion Glycolysis. Clinical positron imaging : official journal of the Institute for Clinical PET* **2**:159-171.
28. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, Chaudhari S, Yang D, Schmitt M, Laforest R, Thorstad W, Deasy JO 2009 Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern recognition* **42**:1162-1171.
29. Yoon HJ, Kim Y, Kim BS 2015 Intratumoral metabolic heterogeneity predicts invasive components in breast ductal carcinoma in situ. *European radiology* **25**:3648-3658.
30. Watabe T, Tatsumi M, Watabe H, Isohashi K, Kato H, Yanagawa M, Shimosegawa E, Hatazawa J 2012 Intratumoral heterogeneity of F-18 FDG uptake differentiates between gastrointestinal stromal tumors and abdominal malignant lymphomas on PET/CT. *Annals of nuclear medicine* **26**:222-227.
31. Nakajo M, Nakajo M, Jinguji M, Fukukura Y, Nakabeppu Y, Tani A, Yoshiura T 2015 The value of intratumoral heterogeneity of (18)F-FDG uptake to differentiate between primary benign and malignant musculoskeletal tumours on PET/CT. *The British journal of radiology* **88**:20150552.
32. Hanaoka K, Hosono M, Tatsumi Y, Ishii K, Im SW, Tsuchiya N, Sakaguchi K, Matsumura I 2015 Heterogeneity of intratumoral (111)In-ibritumomab tiuxetan and (18)F-FDG distribution in association with therapeutic response in radioimmunotherapy for B-cell non-Hodgkin's lymphoma. *EJNMMI research* **5**:10.

33. Kang SR, Song HC, Byun BH, Oh JR, Kim HS, Hong SP, Kwon SY, Chong A, Kim J, Cho SG, Park HJ, Kim YC, Ahn SJ, Min JJ, Bom HS 2014 Intratumoral Metabolic Heterogeneity for Prediction of Disease Progression After Concurrent Chemoradiotherapy in Patients with Inoperable Stage III Non-Small-Cell Lung Cancer. *Nuclear medicine and molecular imaging* **48**:16-25.
34. Tixier F, Vriens D, Cheze-Le Rest C, Hatt M, Disselhorst JA, Oyen WJ, de Geus-Oei LF, Visser EP, Visvikis D 2016 Comparison of Tumor Uptake Heterogeneity Characterization Between Static and Parametric 18F-FDG PET Images in Non-Small Cell Lung Cancer. *J Nucl Med* **57**:1033-1039.
35. van Velden FH, Cheebsumon P, Yaqub M, Smit EF, Hoekstra OS, Lammertsma AA, Boellaard R 2011 Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. *Eur J Nucl Med Mol Imaging* **38**:1636-1647.
36. Chung HH, Kang SY, Ha S, Kim JW, Park NH, Song YS, Cheon GJ 2016 Prognostic value of preoperative intratumoral FDG uptake heterogeneity in early stage uterine cervical cancer. *Journal of gynecologic oncology* **27**:e15.
37. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin J-C, Pieper S, Aerts HJWL 2017 Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* **77**:e104-e107.
38. Presotto L, Bettinardi V, De Bernardi E, Belli ML, Cattaneo GM, Broggi S, Fiorino C 2018 PET textural features stability and pattern discrimination power for radiomics analysis: An "ad-hoc" phantoms study. *Phys Med* **50**:66-74.
39. Zwanenburg A, Leger S, Vallières M, Lock S 2019 Image biomarker standardisation initiative. 2019.
40. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D 2012 Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. *J Nucl Med* **53**:693-700.
41. Orhac F, Soussan M, Maisonobe J-A, Garcia CA, Vanderlinden B, Buvat I 2014 Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med* **55**:414-422.

42. De Lellis R, Lloyd R, Heitz P, Eng C 2004 Pathology and genetics of tumour of endocrine organs. . In: Kleihues P, Sobin L, (eds) WHO Classification of Tumours. Vol. IARC, Lyon, 49-133.
43. Giovanella L, Ceriani L, Treglia G 2014 Role of isotope scan, including positron emission tomography/computed tomography, in nodular goitre. Best practice & research Clinical endocrinology & metabolism **28**:507-518.

Table 1. Demographic and clinicopathological characteristics of patients.

Variable	Patients (n = 104)	Benign lesions (n = 77)	Malignant lesions (n = 30)	P-value***
Sex (Female/Male)*	67/37 (64%/36%)	51/26 (66%/34%)	19/11 (63%/37%)	0.95
Age (years)**	65 (54–74)	65 (54–75)	62.5 (52–72)	0.27
TSH (mUI/L)**	1.5 (0.96–2.10)	1.5 (0.86–2.10)	1.6 (1.2–2.80)	0.15
Largest tumor diameter on US (mm)**	17 (12–26)	17 (12–23.5)	21 (14–30.5)	0.10
Final diagnosis		Benign goiter 62 AFTN 1 AIT 14	Thyroid carcinoma 26 Papillary 19 Follicular 5 Medullary 1 Anaplastic 1 Non-Hodgkin lymphoma 1 Metastatic lesion 3 Sarcoma 1 Renal cancer	

			1	
			Esofageous cancer	
			1	

AAFTN: autonomously functioning thyroid nodules, AIT:

autoimmune thyroiditis, TSH: thyroid stimulating hormone, US:

ultrasound

data expressed as *frequency or **median and interquartile range

*** χ^2 test for comparison of frequencies and Mann-Whitney U-test for comparison of medians

Table 2. Distribution of PET functional parameters and nonredundant radiomics features in the two subgroups of patients with benign and malignant nodular disease.

Variable	Benign lesions (<i>n</i> = 77)	Malignant lesions (<i>n</i> = 30)	<i>P</i> -value*
PET metrics			
SUVmax	6.71 (4.52–10.96)	10.81 (7.21–14.71)	0.0146
SUVmean	2.79 (2.39–3.57)	3.88 (2.78–5.64)	0.0040
SUVpeak	3.45 (2.98–5.57)	7.11 (4.25–10.90)	0.0005
MTV (mL)	3.64 (2.48–7.22)	10.75 (4.10–29.66)	0.0002
TLG	11.61 (8.73–15.30)	44.69 (19.55–81.28)	<0.0001
AUC-CSH	0.42 (0.40–0.50)	0.44 (0.42–0.50)	0.5930
CV	0.40 (0.29–0.59)	0.54 (0.28–0.65)	0.2580
Radiomics			
shape_Sphericity	0.67 (0.53–0.78)	0.79 (0.72–0.82)	0.0004
shape_Maximum2DDiameterSlice	32.20 (23.20–45.30)	22.85 (20.12–29.55)	0.0018
firstorder_Energy	5592.0 (1533.40–20203.00)	1178 (629.30–3541.40)	0.0005
GLCM_Contrast	150.30 (86.50–265.20)	253.8 (179.10–328.10)	0.0025
GLCM_Inverse difference moment	0.13 (0.11–0.18)	0.09 (0.07–0.13)	0.0003
GLCM_SumSquares	183.20 (142.30–226.10)	212.10 (185.80–239.30)	0.0380

PET: positron emission tomography, SUV: standardized uptake value, MTV: metabolic tumor volume, TLG: total lesion glycolysis, AUC-CSH: area under the curve-cumulative SUV volume histograms, CV: coefficient of variation, GLCM: gray-level co-occurrence matrix.

Data expressed as median and interquartile range. *Mann-Whitney U-test.

Table 3. PET metrics and radiomics parameters, ROC curve analysis and univariate logistic regression results.

Variable	AUC	Best Cut-point	P-value	Sensitivity (%)	Specificity (%)	Logistic regression P-value	Odds ratio (95% CI)	NPV (%)	PPV (%)	Accuracy (%)
SUVmax	0.652	>7.05	0.0133	76.7	57.1	0.0025	4.38 (1.68–11.43)	86	41	74
SUVmean	0.68	>3.03	0.0037	50.0	83.1	0.0008	4.92 (1.94–12.5)	81	54	74
SUVpeak	0.717	>3.91	0.0003	80.0	58.4	0.0007	5.62 (2.06–15.33)	88	43	64
MTV	0.733	>8.63	0.0001	60.0	81.8	0.0003	5.29 (2.14–13.12)	83	51	73
TLG	0.7	>39.1	<0.00	56.7	88.3	<0.000	9.88	84	65	79

	56	5	01			1	(3.6 3– 26.9 2)			
AUC-CSH	0.5 33	>0.42	0.580	66.7	50.6	0.1096	2.05 (0.8 5– 4.95)	80	34	55
CV	0.5 71	>0.59	0.283	43.3	76.6	0.0441	2.51 (1.0 2– 6.13)	78	42	67
shape_Spheri city	0.7 23	<57.9 9	0.000 4	46.7	96.1	<0.000 1	21.5 8 (5.5 4– 84.0 1)	82	82	82
shape_Maxi mum2D DiameterSlice	0.6 95	>31.3 2	0.001 8	53.3	84.4	0.0005	5.42 (2.1 0– 13.9 2)	81	55	75
firstorder_En ergy	0.7 16	>1177 .97	0.000 3	86.7	50.6	0.0011	6.67 (2.1 3– 20.9 3)	91	41	61

GLCM_Contra st	0.6 99	<163. 53	0.001 3	56.7	84.4	0.0002	6.19 (2.4 0– 15.9 3)	82	57	76
GLCM_Invers e Difference Moment	0.7 28	>10	<0.00 01	83.3	55.8	0.0007	6.32 (2.1 9– 18.2 5)	89	42	64
GLCM_SumSq uares	0.6 29	<177. 59	0.052 6	46.7	83.1	0.0021	4.31 (1.6 9– 10.9 4)	80	52	73

PET: positron emission tomography, ROC: receiver operator characteristic, SUV: standardized uptake value, MTV: metabolic tumor volume, TLG: total lesion glycolysis, AUC-CSH: area under the curve-cumulative SUV volume histograms, CV: coefficient of variation, GLCM: gray-level co-occurrence matrix, CI: confidence interval, NPV: negative predictive value, PPV: positive predictive value

Table 4. Multivariate analysis (stepwise logistic regression) results.

Variable	Odds ratio (95% CI)	P-value	Likelihood ratio test	Hosmer and Lemeshow goodness of fit test	Bootstrap ($R = 1000$) optimism-corrected AUC-ROC
SUVmax	4.15 (1.22–14.19)	0.0231	<0.0001	0.9999	0.830
TLG	4.52 (1.39–14.71)	0.0123			
shape_Sphericity	18.11 (3.68–89.11)	0.0004			

PET: positron emission tomography, AUC-ROC: area under the curve-receiver operator characteristic, SUV: standardized uptake value, TLG: total lesion glycolysis, CI: confidence interval.

Figure Legends

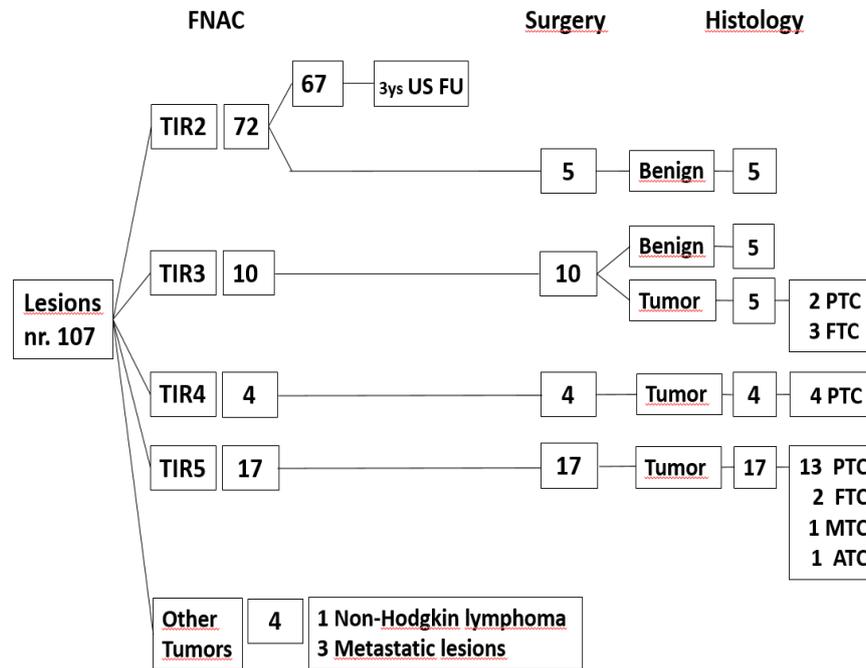


Figure 1. Patient evaluations and results of diagnostic work-up.

PTC: Papillary Thyroid Carcinoma; FTC: Follicular Thyroid Carcinoma, MTC: Medullary Thyroid Carcinoma, ATC: Anaplastic Thyroid Carcinoma; US FU: ultrasounds follow-up; FNAC: fine-needle aspiration cytology, TIR: FNAC results reported according to the Italian Consensus for the Classification and Reporting of Thyroid Cytology [TIR1 (inadequate), TIR2 (benign), TIR3 (indeterminate), TIR4 (suspicious for malignancy), TIR5 (malignancy)].