# 18F-FDG-Based Radiomics and Machine Learning

Thomas Godefroy, Gauthier Frécon, Antoine Asquier-Khati, Diana Mateus, Raphaël Lecomte, Mira Rizkallah, Nicolas Piriou, Bastien Jamet, Thierry Le Tourneau, Amandine Pallardy, et al.

JACC: CARDIOVASCULAR IMAGING VOL. 16, NO. 7, 2023

PUBLISHED BY ELSEVIER

ORIGINAL RESEARCH

# $^{18}$F-FDG-Based Radiomics and Machine Learning

## Useful Help for Aortic Prosthetic Valve Infective Endocarditis Diagnosis?

Thomas Godefroy, MD,[a] Gauthier Frécon, MSc,[a,c] Antoine Asquier-Khati, MD,[b] Diana Mateus, PhD,[c] Raphaël Lecomte, MD,[b] Mira Rizkallah, PhD,[c] Nicolas Piriou, MD,[a,d] Bastien Jamet, MD,[a] Thierry Le Tourneau, MD, PhD,[d] Amandine Pallardy, MD,[a] David Boutoille, MD, PhD,[b] Thomas Eugène, MD,[a] Thomas Carlier, PhD[a]

ABSTRACT

**BACKGROUND** Fluorine-18 fluorodeoxyglucose ($^{18}$F-FDG)-positron emission tomography (PET)/computed tomography (CT) results in better sensitivity for prosthetic valve endocarditis (PVE) diagnosis, but visual image analysis results in relatively weak specificity and significant interobserver variability.

**OBJECTIVES** The primary objective of this study was to evaluate the performance of a radiomics and machine learning–based analysis of $^{18}$F-FDG PET/CT (PET-ML) as a major criterion for the European Society of Cardiology score using machine learning as a major imaging criterion (ESC-ML) in PVE diagnosis. The secondary objective was to assess performance of PET-ML as a standalone examination.

**METHODS** All $^{18}$F-FDG-PET/CT scans performed for suspected aortic PVE at a single center from 2015 to 2021 were retrospectively included. The gold standard was expert consensus after at least 3 months' follow-up. The machine learning (ML) method consisted of manually segmenting each prosthetic valve, extracting 31 radiomics features from the segmented region, and training a ridge logistic regressor to predict PVE. Training and hyperparameter tuning were done with a cross-validation approach, followed by an evaluation on an independent test database.

**RESULTS** A total of 108 patients were included, regardless of myocardial uptake, and were divided into training (n = 68) and test (n = 40) cohorts. Considering the latter, PET-ML findings were positive for 13 of 22 definite PVE cases and 3 of 18 rejected PVE cases (59% sensitivity, 83% specificity), thus leading to an ESC-ML sensitivity of 72% and a specificity of 83%.

**CONCLUSIONS** The use of ML for analyzing $^{18}$F-FDG-PET/CT images in PVE diagnosis was feasible and beneficial, particularly when ML was included in the ESC 2015 criteria. Despite some limitations and the need for future developments, this approach seems promising to optimize the role of $^{18}$F-FDG PET/CT in PVE diagnosis.
(J Am Coll Cardiol Img 2023;16:951-961) © 2023 by the American College of Cardiology Foundation.

From the [a]Nantes Université, CHU Nantes, INSERM, Nuclear Médicine, Nantes, France; [b]Nantes Université, CHU Nantes, INSERM, Infectious Diseases Department, Nantes, France; [c]ECN, LS2N, Nantes, France; and the [d]Nantes Université, CHU Nantes, CNRS, INSERM, l'institut du Thorax, Nantes, France.
The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received December 9, 2022; accepted January 25, 2023.

ISSN 1936-878X/$36.00 https://doi.org/10.1016/j.jcmg.2023.01.020

**ABBREVIATIONS AND ACRONYMS**

**$^{18}$F-FDG** = fluorine-18 fluorodeoxyglucose

**CT** = computed tomography

**ML** = machine learning

**PET** = positron emission tomography

**PVE** = prosthetic valve endocarditis

**ROI** = region of interest

**SUV** = standardized uptake value

Following the publications of the European Society of Cardiology (ESC) 2015 criteria,[1] and the American College of Cardiology/American Heart Association guidelines in 2020,[2] fluorine-18 fluorodeoxyglucose ($^{18}$F-FDG) positron emission tomography (PET)/computed tomography (CT) became a routinely performed examination for patients with suspected prosthetic valve infective endocarditis (PVE), thereby increasing the diagnostic sensitivity when added to the Duke criteria.[3-5]

Despite acceptable diagnostic performances in this setting, $^{18}$F-FDG PET/CT lacks standardized visual and quantitative criteria for interpretation, with multiple confounding factors explaining the relatively moderate interobserver reproducibility observed.[6] Although conventional quantitative measurements of fluorodeoxyglucose (FDG) uptake may be useful in the setting of PVE diagnosis,[7,8] their ability to assess FDG distribution within a region of interest (ROI) and from significant variations among centers is weak, thus making their routine application difficult.[9] Consequently, scan interpretation remains almost exclusively based on human qualitative visual assessment.

The advent of artificial intelligence algorithms in cardiovascular imaging opened new perspectives in terms of image preprocessing, analysis and interpretation. Machine learning (ML) algorithms, at the core of artificial intelligence processes, perform classification or risk prediction tasks after observing and recognizing patterns in broad-range data they have been trained on. Such algorithms are therefore particularly suited to the field of nuclear imaging and are capable of taking into account and associating multiple data, visible or not (radiomics), to reproduce a version of truth, thereby potentially enhancing human interpretation. Although ML-based algorithms seem to be promising for the diagnosis of ischemic cardiopathy,[10] scarce published reports are available on infective endocarditis.

We therefore aimed to evaluate the potential ability of quantitative analysis of $^{18}$F-FDG PET/CT scans using radiomics and ML for PVE diagnosis.

## METHODS

**PATIENTS.** Between 2015 and 2021, all patients who had exclusively an aortic prosthetic heart valve and who underwent $^{18}$F-FDG PET/CT for suspected PVE in the Department of Nuclear Medicine of the University Hospital of Nantes (Nantes, France) were retrospectively included. Clinical, biologic, microbiologic, and imaging data (transthoracic and transesophageal echocardiography, $^{18}$F-FDG PET/CT scans) were collected on admission and until 3-month follow-up. The data set was divided into a training set, including all examinations acquired between 2015 and 2019, and a test set, with examinations acquired between 2019 and 2021. The Institutional Review Board approved this study, and written informed consent was obtained from all the participating patients.

**PET/CT IMAGING PROTOCOL AND VISUAL ANALYSIS.** Twelve-hour fasting and a special low-carbohydrate diet regimen the day before were required to reduce the physiologic uptake of the myocardium. All $^{18}$F-FDG PET/CT images were acquired in the same center using 2 different systems. Whole-body $^{18}$F-FDG PET/CT was carried out after injection of 3 MBq/kg of $^{18}$F-FDG. Patients for the training cohort were all acquired using a Siemens Biograph mCT (Siemens Healthcare), whereas all patients in the testing cohort were acquired using a Siemens Biograph Vision 450. Data were reconstructed using the 3-dimensional ordinary Poisson ordered subset expectation maximization algorithm with point spread function recovery and time-of-flight information. The reconstruction parameters were as follows: 3 iterations, 21 subsets with a 2-mm full width at half-maximum (FWHM) Gaussian post-filtering for the Biograph mCT (voxel size $4 \times 4 \times 2\,mm^3$); and 4 iterations, 5 subsets with a 3-mm FWHM Gaussian post-filtering for the Biograph Vision 450 (voxel size $1.65 \times 1.65 \times 1.65\,mm^3$). Images were reconstructed in standardized uptake value (SUV) normalized for body weight. Low-dose 3-dimensional CT was acquired for attenuation and scatter correction, without intravenous contrast enhancement, and automatic mA according to the patient's weight.

All $^{18}$F-FDG PET/CT scans were analyzed, regardless of initiation of antibiotic therapy or myocardial uptake. A visual qualitative analysis of the myocardial extinction quality and the valvular prosthesis uptake was performed independently by 2 nuclear medicine physicians (1 expert and 1 nonexpert with, respectively, 9 years and 1 year of clinical cardiac PET/CT imaging experience) blinded to the clinical, biologic, and imaging data. Images were interpreted and reoriented in the prosthetic aortic valve plane on a dedicated workstation (Syngo.via, Siemens Healthcare). $^{18}$F-FDG PET/CT results were considered positive when the valvular uptake pattern was heterogeneous or extending beyond the perivalvular area. In contrast, an absence or homogenous uptake of the prosthetic valve were considered normal. An

additional analysis of the whole-body acquisition was subsequently performed and visually analyzed by both nuclear medicine physicians, to identify potential septic embolic events and abnormal splenic or bone marrow uptake.

**SEGMENTATION, FEATURES PROCESSING, AND SELECTION.** The aortic prosthetic valve ROI was manually segmented in 3 dimensions on $^{18}$F-FDG PET images excluding the myocardium uptake. Radiomics features (excluding all shape features because they are meaningless in this context) were computed on this segmented volume using PyRadiomics software version 3.0 (PyRadiomics).[11] Before computation, $^{18}$F-FDG PET images were resampled to the same voxel size ($2 \times 2 \times 2$ mm$^3$) using a bicubic spline interpolation. Data were subsequently normalized with a linear equalization using 64 bins. Among features available for computation, a subset of 31 features was preselected (Supplemental Figure 1), choosing those that presented the best properties of repeatability.[12-14] The reliability of textural radiomics features (excluding first order metrics) was subsequently derived following the methodology proposed by Pfaehler et al.[15] Briefly, for each patient and each radiomics feature, the signal within the ROI was randomly shuffled 50 times, and the radiomics feature was subsequently computed. The radiomics feature was considered unreliable if its original value lay within the 95% CI computed using the random shuffling. Finally, the radiomics feature was retained if its original value was outside the proposed CI for 90% of the patients.

Because many radiomics features (including first order features) could be highly correlated, a variation inflation factor (using a threshold of 100) was applied to retain only uncorrelated features in the model.

In short, the PET images were manually segmented in 3 dimensions to form a volume of interest limited to the valve region. The volume of these regions of interest was variable among patients, but all images were resampled and normalized for analysis at a same voxel size ($2 \times 2 \times 2$ mm$^3$).

Additionally, we considered 5 clinically based features related to $^{18}$F-FDG PET/CT imaging, hereafter named extraprosthetic features. The first 2 features were elapsed time between PET/CT imaging and initiation of antibiotics and the elapsed time between $^{18}$F-FDG PET/CT imaging and prosthetic surgery; both features are known to influence valvular uptake of $^{18}$F-FDG PET/CT. The remaining 3 features were parameters extracted from the visual analysis of whole-body $^{18}$F-FDG PET/CT: septic embolic events outside the cardiac area; spleen uptake (defined as positive when greater than the visual liver uptake); and bone marrow medullary uptake (defined as positive when greater than the visual liver activity). All features were then standardized setting the mean to 0 and the SD to 1 to make the features space homogeneous.

**MACHINE LEARNING MODEL.** The objective of the model was to classify the presence or absence of PVE correctly. As such, data augmentation on the training set (imbalance 63%/37% in favor of the positive class) was used to increase the minority class cardinal to the one of the majority class. Both the synthetic minority oversampling technique[16] and the adaptive synthetic sampling[17] approaches for imbalanced learning were considered. A logistic regression including ridge regularization was considered the ML model. A probability cutoff of 0.5 was used to classify a patient as having confirmed or rejected PVE.

The training data set included all patients acquired using the Biograph mCT PET/CT system between 2015 and 2019. The external testing data set included patients from 2019 to 2021 acquired using a different PET system from the testing data set, namely, the Biograph Vision 450 PET/CT system. A leave-1-out cross validation on the training set was performed to select the hyperparameters including the method used for data augmentation (Supplemental Table 1). For each set of parameters, models were evaluated 3 times by leave-1-out cross validation, and the hyperparameters that led the best balanced accuracy were chosen. Once the best hyperparameters were chosen, the training was repeated 100 times on the whole training data set and each time evaluated using the test data set. An evaluation that led to the median balanced accuracy among the 100 loops was reported as the final prediction of the algorithm. Additionally, an end-to-end deep learning approach was implemented for comparison with the aforementioned model. The deep learning model is presented in the Supplemental Methods.

**ENDPOINTS AND STATISTICAL ANALYSIS.** The gold standard for the diagnosis of definitive PVE consisted of expert consensus among the endocarditis team established at least 3 months after admission and based on data obtained during follow-up. Such data included results of clinical findings, microbiologic findings, repeat imaging of the new information, visual and pathologic evaluation of explanted prostheses, and molecular biology techniques. The primary endpoint was to evaluate sensitivity, specificity, and balanced accuracy of the ESC 2015 criteria by using our ML-based algorithm result to define the $^{18}$F-FDG PET/CT major criteria (ESC-ML). Diagnostic performances of the conventional ESC 2015 criteria,

**TABLE 1 Comparison of Demographic and Clinical Data Between Training and Test Data Sets**

| | Train/Validation (n = 68) | Test (n = 40) | *P* Value |
|---|---|---|---|
| Demographic and clinical data | | | |
| Age, y | 75 (67.0-82.0) | 80 (70.75-84.25) | 0.05 |
| Male | 52 (76.5) | 31 (77.5) | 1 |
| Charlson comorbidity index | 5 (3.5-6.5) | 5 (5.0-7.25) | 0.95 |
| Transcatheter aortic valve replacement | 14 (20.6) | 12 (30.0) | 0.36 |
| Bioprosthetic valve | 41 (60.3) | 23 (57.5) | 0.84 |
| Mechanical valve | 12 (17.6) | 5 (12.5) | 0.43 |
| Clinical presentation | | | |
| Fever | 47 (69.1) | 34 (85.0) | 0.07 |
| C-reactive protein, mg/L | 47 (13.3-101.75) | 90 (57.5-142.0) | **0.01** |
| Blood cultures positive for IE (Duke major criteria) | 49 (72) | 25 (62.5) | 0.39 |
| Echocardiography positive for IE (Duke major criteria) | 24 (35.3) | 14 (35) | 1 |
| Time between PET/CT and prosthetic valve implantation, mo | 48 (24-96) | 60 (24.0-99.0) | 0.96 |
| Time between PET/CT and initiation of antibiotics, d | 13 (7.43-21.48) | 12 (9.0-16.0) | 0.539 |

Values are mean (range) or n (%). **Bold** indicates a significant difference.
CT = computed tomography; IE = infective endocarditis; PET = positron emission tomography.

including qualitative visual interpretation of the $^{18}$F-FDG PET/CT scans by the nuclear medicine expert who reached the best performance assessment regarding balanced accuracy with respect to endocarditis team consensus (ESC-visual), as well as the modified Duke criteria were also assessed. Three secondary endpoints were defined: 1) sensitivity, specificity, and balanced accuracy of the machine learning-based analysis of $^{18}$F-FDG PET/CT (PET-ML) exclusively considering the aortic valve; 2) sensitivity, specificity, and balanced accuracy of our model using the PET signal supplemented by the clinical features detailed previously (PET/clinical-ML); and 3) sensitivity, specificity, and balanced accuracy of the ESC 2015 criteria by combining the human visual analysis and the result of our ML-based algorithm (ie, examination results were regarded as positive only if the algorithm and the human interpreter were in agreement; all other cases were considered negative). This new criterion was named ESC-combined. For primary and secondary endpoints, our model performance was compared with the conclusion of the visual analysis of the nuclear medicine expert who reached the best performance conclusion regarding balanced accuracy and with respect to endocarditis team consensus (PET-visual). The receiver-operating characteristic curve and the associated area under the curve corresponding to the median among the 100 loops were also reported for the PET-ML model.

The comparison among sensitivity, specificity, and accuracy of the different approaches was performed using the McNemar test. Interobserver reproducibility between the 2 nuclear medicine experts was measured by calculating the Cohen kappa coefficient. A Wilcoxon or Fisher test was used to compare patients' clinical characteristics. A value of $P < 0.05$ was considered significant. All statistical analysis, feature processing, and algorithms were computed using Python 3.9. The Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation checklist for ML[18] is provided in Supplemental Table 2.

## RESULTS

**PATIENT CHARACTERISTICS.** Of the 108 consecutive patients included in the study, 65 were classified as having confirmed PVE and 43 as rejected PVE as defined by an expert consensus of the endocarditis team after at least 3 months of follow-up as the gold standard. Characteristics and management of the 108 patients are presented in Supplemental Table 3. The median age of our cohort was 77 years (IQR: 68-82 years), and most patients were male (76.9%). The median time between $^{18}$F-FDG PET/CT and prosthetic surgery was 60 months (IQR: 24-96 months). Blood culture results were positive for 90.8% of the patients with confirmed PVE, and the most common species were streptococci (45.8%), enterococci (20.3%), and *Staphylococcus aureus* (17%). According to the Duke criteria, 38 patients were identified as having definite PVE (35.2%), 55 as possible PVE (50.9%), and 15 as rejected PVE (13.9%). The median duration between $^{18}$F-FDG PET/CT and blood culture positivity was 15 days (range: 11-21 days), whereas the median time between initiation of intravenous antibiotics and $^{18}$F-FDG PET/CT was 13 days (range: 7.54-19 days). The training set consisted of 68 patients (63%), and the test set included 40 patients (37%). Both the training and test groups were balanced in terms of demographic, clinical, biologic, and imaging data, except for the C-reactive protein analysis, which was significantly higher in the test set (C-reactive protein, 90 mg/L vs 47 mg/L; $P = 0.01$). A comparison between the 2 groups is presented in **Table 1**.

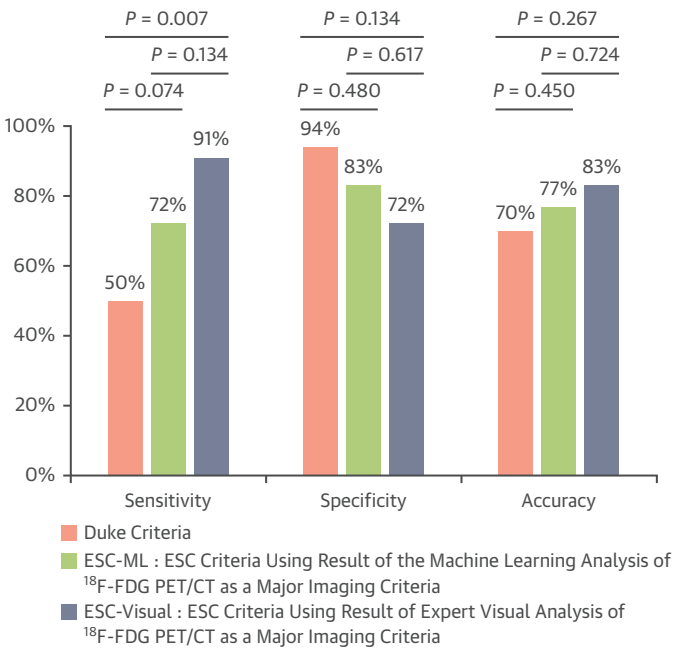
**PRIMARY ENDPOINT.** Sensitivity, specificity, and balanced accuracy of the ESC-ML were 72%, 83%, and 77% respectively, as opposed to the ESC-visual criteria diagnostic performance with expert visual assessment, which were 91%, 72%, and 82% (**Figure 1, Central Illustration**). No significant differences were observed between metrics of ESC-ML and ESC-visual ($P = 0.13$, $P = 0.61$, and $P = 0.72$, respectively).

Sensitivity, specificity, and accuracy of the Duke criteria were 50%, 94%, and 70% respectively. The ESC-ML reclassified 7 patients considered as having "possible PVE" according to the modified Duke criteria into the "definitive PVE" category, among which 5 true positive and 2 false negative results were observed, thus explaining a higher sensitivity than that observed with the modified Duke criteria (**Figure 2**). No significant difference was observed between the Duke criteria and ESC-ML in terms of sensitivity, specificity, and balanced accuracy ($P = 0.07$, $P = 0.47$, and $P = 0.45$, respectively). Sensitivity of the ESC-visual was significantly higher than that achieved with the Duke criteria ($P = 0.007$) but specificity and accuracy were not significantly different ($P = 0.13$ and $P = 0.26$, respectively). The ESC-visual sensitivity, specificity, and balanced accuracy that were based on the nonexpert physician's visual assessment were, respectively, 68%, 72%, and 70%. The best parameters retained for the model with and without extraprosthetic features are provided in Supplemental Tables 4 and 5.

**SECONDARY ENDPOINTS.** Considering exclusively the aortic prosthetic valve uptake, the PET-ML was positive for 13 of 22 patients with confirmed PVE, and it was negative for 15 of 18 patients with rejected PVE (59 % sensitivity, 83% specificity, 70% balanced accuracy). Sensitivity, specificity, and balanced accuracy that were based on visual assessment of PET (PET-visual) of the expert and nonexpert nuclear medicine physicians were, respectively, 77% vs 50 %, 72% vs 72%, and 75% vs 60%. Interobserver reproducibility was low, with a kappa coefficient of 0.51. The expert nuclear medicine review concluded with a positive $^{18}$F-FDG PET/CT result for 15 of 22 patients with confirmed PVE and with a negative $^{18}$F-FDG PET/CT result for 13 of 18 patients with rejected PVE (77% sensitivity, 72% specificity, 75% balanced accuracy). No significant difference was reported ($P = 0.77$) between PET-ML and the expert visual analysis of $^{18}$F-FDG-PET/CT (**Figure 3**).

Adding clinical features related to $^{18}$F-FDG PET/CT into the model (PET/clinical-ML) did not modify the diagnostic performances of either PET-ML or ESC-ML. The sensitivity, specificity, and balanced accuracy of ESC-combined were 72%, 89%, and 80% for the expert and 68%, 89%, and 79% for the nonexpert, respectively. The receiver-operating characteristic curve (median over the 100 loops) for the PET-ML model is given in **Figure 4** and led to an area under the curve of 0.82. Results obtained using the deep learning-based approach exhibited moderate performance when compared with the PET-ML. These results are presented in the Supplemental Table 6.

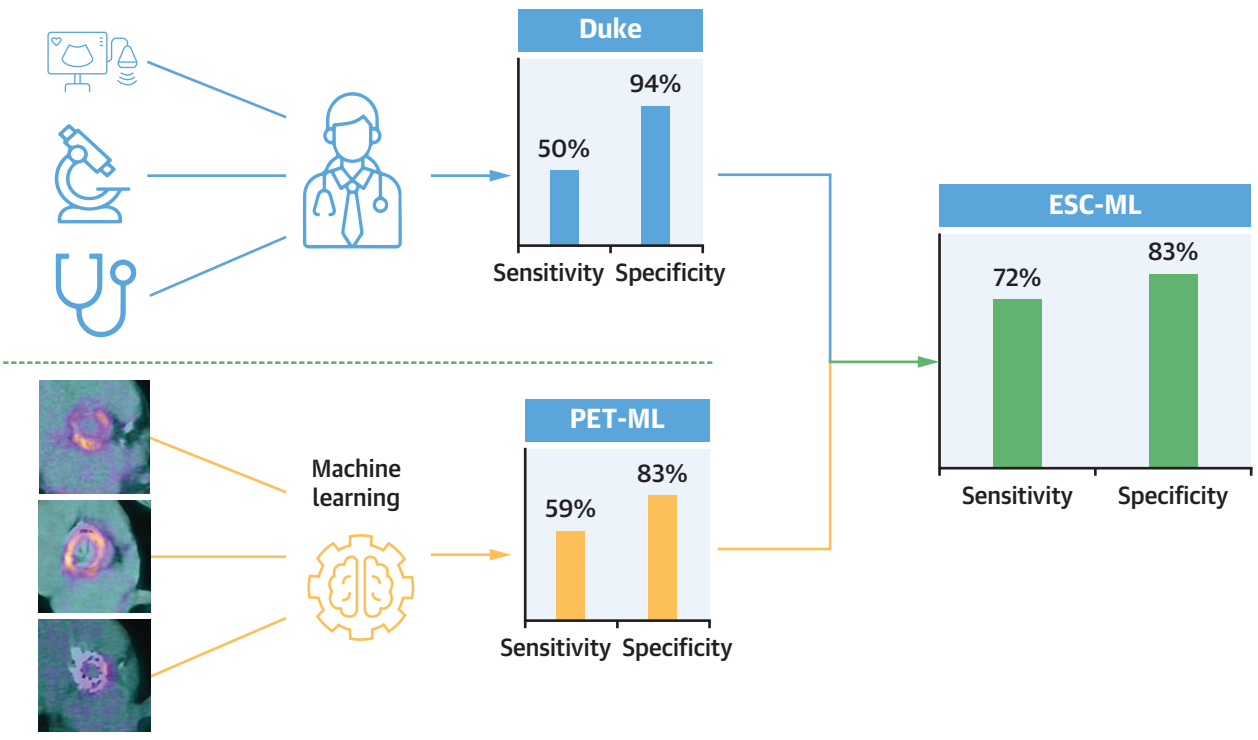
**FIGURE 1 Diagnosis Performance of Modified Duke Criteria vs ESC-ML Criteria and ESC-Visual Criteria**




Both European Society of Cardiology (ESC) 2015 score using $^{18}$F-FDG PET/CT with machine learning as a major imaging criterion (ESC-ML) and European Society of Cardiology 2015 score using PET-visual as the major imaging criterion (ESC-visual) showed a trend to improve sensitivity over Duke criteria, but only the ESC-visual was significantly different. $^{18}$F-FDG = fluorine-18 fluorodeoxyglucose; CT = computed tomography; PET = positron emission tomography.

**FEATURE IMPORTANCE.** It is worthwhile to highlight the importance of each feature for the classification task. For logistic regression, predictions are calculated from the weighted sum of feature inputs. The absolute values of those weights provide a good indication of feature importance. Moreover, a positive weight sign tends to classify patients in the positive class when the feature value increases. Inversely, a negative weight sign tends to classify patients in the negative class when the feature value increases. A logistic regression model was trained 100 times on the concatenation of the initial training and test sets by using the same hyperparameters chosen by our leave-1-out cross validation. The features' importance was then extracted over the 100 trainings (**Figure 5**). Several first order textural and extraprosthetic features seemed to be important for the model. First Order Mean, First Order Energy, Gray Level Co-occurrence Matrix (GLCM) ClusterShade, Gray Level Dependence Matrix (GLDM)

**CENTRAL ILLUSTRATION** **Process and Performance of the European Society of Cardiology 2015 Criteria Using Our Machine Learning-Based Algorithm Result as a Major Imaging Criterion**




Godefroy T, et al. J Am Coll Cardiol Img. 2023;16(7):951-961.

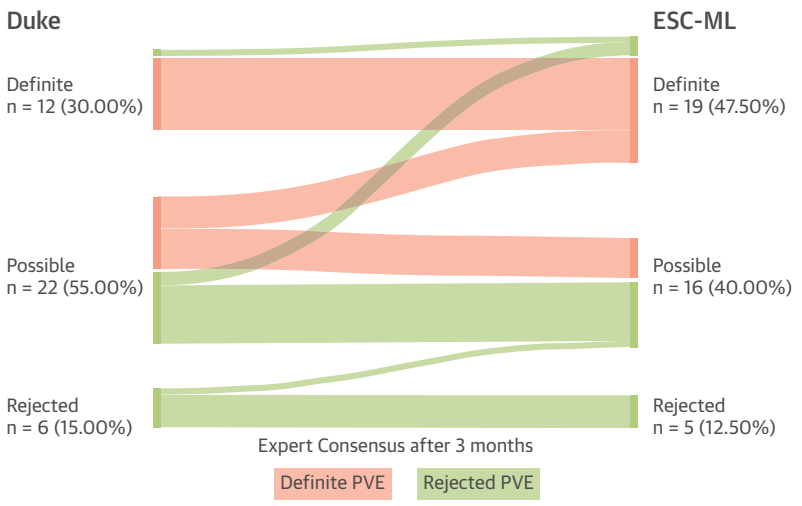
Addition of our machine learning analysis of $^{18}$F-FDG PET/CT (PET-ML) to the modified Duke criteria increases the sensitivity of Duke criteria but decreases specificity. $^{18}$F-FDG = fluorine-18 fluorodeoxyglucose; CT = computed tomography; PET = positron emission tomography; PET-ML = machine learning-based analysis of $^{18}$F-FDG PET/CT.

SmallDependenceHighGrayLevelEmphasis, as well as extraprosthetic features (except the elapsed time between PET/CT imaging and prosthetic surgery), were found to be valuable parameters for the model performance. The correlation of all features of the model with the outcome is provided in Supplemental Figure 2 and Supplemental Table 7.

## DISCUSSION

To our knowledge, this is the first study to investigate the potential value of ML algorithms in the diagnosis of PVE. Considering only $^{18}$F-FDG PET/CT images (PET-ML), the ML model yielded performances that were similar in terms of sensitivity and specificity to those in the published reports (ranges: 73%-100% and 71%-100%, respectively) with visual interpretations from experts.[19] The integration of PET-ML results to the ESC criteria (ESC-ML) reached acceptable diagnostic performance in our study, and it did not differ significantly from the conventional ESC-visual analysis. The results obtained were similar in magnitude to those presented by Philip et al[6] in a prospective study that assessed sensitivity and specificity of ESC criteria for the diagnosis of PVE, with visual interpretation of the $^{18}$F-FDG PET/CT results (sensitivity, 72%; specificity, 83%). Of note, the sensitivity of ESC-ML was slightly lower than that of ESC-visual, and it did not differ significantly from the Duke criteria sensitivity, although the *P* value nearly achieved the significance threshold ($P = 0.07$). In contrast, specificity of the ESC-ML was higher and interestingly did not differ significantly from the Duke criteria, which are known to be highly specific.[20] The combination of visual and automatic analysis (ESC-combined) showed a balanced accuracy similar to that of the expert ESC-visual (80% vs 82%), but it markedly improved the specificity of the nonexpert ESC-visual (89% vs 72%) without altering sensitivity

**FIGURE 2** **Reclassification of Patients From Duke Modified Criteria to ESC-ML by Adding Machine Learning Analysis of $^{18}$F-FDG PET/CT as a Major Imaging Criterion**



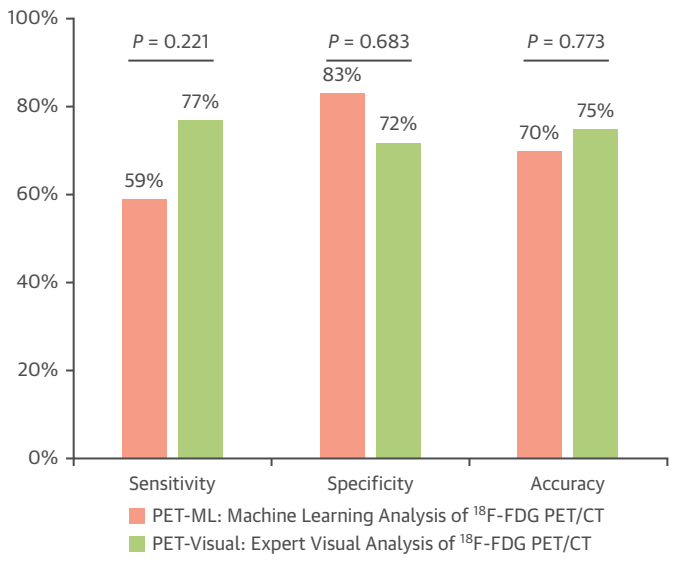

Adding machine learning analysis of $^{18}$F-FDG PET/CT to the European Society of Cardiology (ESC) 2015 criteria reclassified 7 patients considered as having "possible prosthetic valve endocarditis (PVE)" according to the modified Duke criteria into "definitive prosthetic valve endocarditis," among which 5 true positive and 2 false negative results were observed. Abbreviations as in Figure 1.

(68% vs 68%), thus leading to a better balanced accuracy (79% vs 70%) close to that of the expert.

$^{18}$F-FDG is not a specific tracer. It is taken up by all activated immune cells and therefore does not allow an easy distinction between inflammation and infection. The frequent and highly fluctuating tracer uptake on noninfected valves makes visual interpretation very challenging in clinical routine assessment.[21] Better performance of the expert nuclear physician compared with the nonexpert physician was observed. Although this finding suggests that interpretation requires an experienced reader, the final diagnosis remains subject to significant variability of interobserver reproducibility.[6] To overcome this problem, the development of new tracers specific to infectious processes is an option that has been explored by several teams.[21] However, the development of such probes remains uncertain, and the wide diffusion of $^{18}$F-FDG has required performance improvements, specifically, through the development of quantitative analysis tools. Although some studies have suggested a potential benefit of semi-quantitative parameters such as SUV, a meta-analysis highlighted significant variations of thresholds among centers.[9] Therefore, the SUV threshold may not be applicable to other centers.

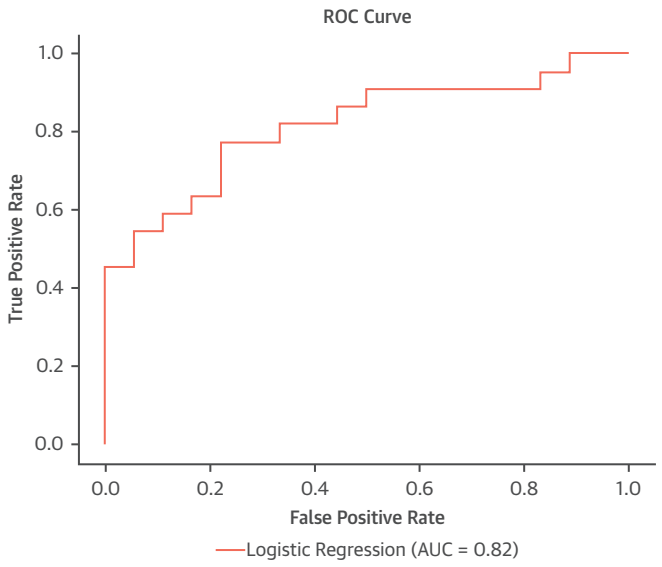
Furthermore, SUV has a limited ability to assess the heterogeneity of FDG distribution within the ROI.

**FIGURE 3** **Diagnosis Performance of PET-ML vs PET-Visual**




When only considering the valve images, machine learning (result of the machine learning analysis of $^{18}$F-FDG PET/CT [PET-ML]) and expert visual analysis (result of qualitative visual interpretation of the $^{18}$F-FDG PET/CT by the nuclear medicine expert [PET-visual]) performance were of the same order of magnitude, without significant difference. Abbreviations as in Figure 1.

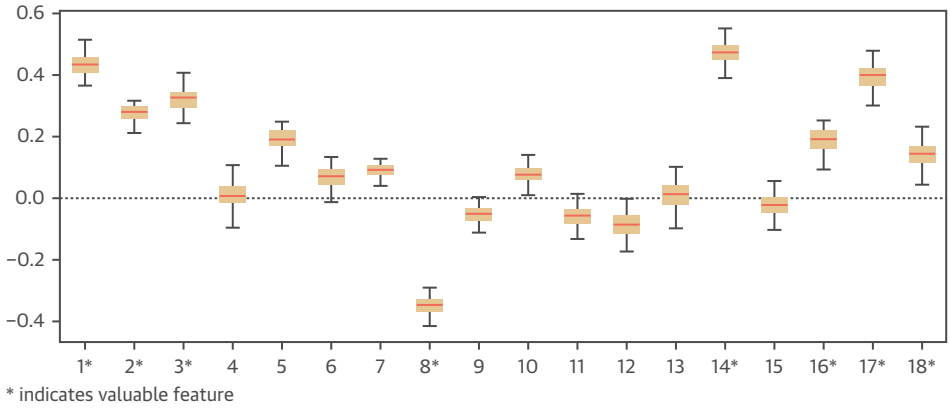
**FIGURE 4 ROC Curve of the PET-ML Model With Clinical Features**




The median receiver-operating characteristic (ROC) curve over the 100 loops reached an area under the curve (AUC) of 0.82 when considering the PET-ML model including the clinical features. Abbreviation as in Figure 3.

Recently, Roque et al[8] introduced the SUV-based "valve uptake index," a semiquantitative parameter that aimed to reflect a degree of homogeneity of the metabolic activity, thereby attenuating the effect of high SUV values when the prosthetic valve uptake is diffuse.[8] In our study, ML results were also guided by the maximum and average uptake values within the ROI, but, as illustrated in Figure 3, the ML also took into account second-order quantitative features (GLCM ClusterShade, GLDM SmallDependenceHighGrayLevelEmphasis, Gray Level Run Length Matrix (GLRM) GrayLevelNonUniformity) that more accurately reflect the heterogeneity of FDG uptake. In the study by Roque et al,[8] a valve uptake index cutoff of >0.45 yielded 85% sensitivity and 88% specificity to confirm PVE, but in the same study, the qualitative "visually abnormal pattern" (focal or heterogenous PV uptake) performed even better, with a sensitivity of 96.7% and a specificity of 96.3%. Importantly, these very good results were, as discussed by these investigators,[8] obtained on selected patients with a clearly classifiable FDG uptake pattern, thus potentially enhancing both the quantitative and visual diagnostic performance. Moreover, all previously published studies excluded patients with suboptimal myocardial extinction (60 of 175 patients excluded in the Endopet study[6]), whereas in our study, all patients were included regardless of myocardial extinction and could therefore be closer to "real-life" conditions. In this work, both nuclear physicians classified myocardial preparation as suboptimal in 12 of 40 patients. In this group of patients, the specificity of both physicians was excellent (100%), but if the sensitivity of the expert physician remained good, that of the nonexpert physician was clearly impaired (83% vs 33%). The performance of the model for these patients was comparatively more balanced, with a sensitivity of 67% and a specificity of

**FIGURE 5 Features Importance According to Logistic Regression**




First Order Mean **(1)**, First Order Energy **(2)**, GLCM ClusterShade **(3)**, GLDM SmallDependenceHighGrayLevelEmphasis **(8)**, Elapsed time between PET/CT imaging and initiation of antibiotics **(14)**, Presence of $^{18}$F-FDG PET/CT septic embolic events outside the cardiac area **(16)**, spleen uptake visually greater than the liver uptake **(17)**, and bone marrow uptake visually greater than the liver uptake **(18)** were found to be valuable parameters for the model performance. Details and correlation of all features of the model with the outcome are provided in Supplemental Figure 2 and Supplemental Table 7. GLCM = Gray Level Co-occurrence Matrix; GLDM = Gray Level Dependence Matrix; other abbreviations as in Figure 1.

67%, suggesting a potential contribution to nonexpert physician's analysis for these difficult cases.

It is well known that the extraction of radiomics features may be influenced by the technical properties of the PET imaging system and the associated reconstruction parameters. In our study, we focused on radiomics features that were proved to be reproducible in the context of multicentric studies. Moreover, validation and training sets were performed on 2 distinct PET imaging systems, providing very similar results (Supplemental Tables 8 and 9) and suggesting possible extrapolation of our results to other systems.

Despite the lack of impact found for extraprosthetic features on ESC-ML diagnostic performance in our study, nearly all of these parameters were found to be valuable for model performance. Diffuse splenic uptake and bone marrow uptake were previously described in the published reports as potential indirect signs of underlying endocarditis,[6,22] thereby corroborating their significant weights observed in the algorithm decision. Of note, the delay between $^{18}$F-FDG PET/CT imaging and prosthetic surgery did not influence the final algorithm decision. This result is in agreement with a previously published study that found no significant impact of this delay on the $^{18}$F-FDG PET/CT diagnostic performance.[4] Elapsed time between initiation of antibiotic therapy and $^{18}$F-FDG PET/CT was the feature with the strongest weight and tended to be positively correlated with the PVE diagnosis. Considering the test group, patients with a positive $^{18}$F-FDG PET/CT result according to the ML model alone tended to have a higher duration of antibiotic therapy before the $^{18}$F-FDG PET/CT scan than did patients with negative examination findings (median: 16 days vs 10 days). The same magnitude of results was observed with the 22 cases of definite PVE, between true positive and false negative subgroups (duration of antibiotics before the $^{18}$F-FDG PET/CT scan: 16 days vs 10 days, respectively). This result is quite surprising because it was thought, and suggested by different studies,[7,23] that prolonged antibiotic therapy led to false negative interpretations and reduced sensitivity as a result of decreased inflammatory activity at the time of $^{18}$F-FDG PET/CT acquisition. This correlation could perhaps be explained by the retrospective design of our study and the fact that $^{18}$F-FDG PET/CT is a second-line examination in clinical routine. Antibiotic therapy is introduced promptly for patients with a strong clinical suspicion of PVE, whereas this therapy may be introduced later in a more doubtful situation. Thus, the pretest probability of patients receiving antibiotic therapy for a longer time may be higher than others. In any case, this finding tends to show that the duration of antibiotic therapy has little influence on the performance of $^{18}$F-FDG PET/CT. Interestingly, 6 of 9 patients with false negative results had a history of transcatheter aortic valve replacement. To our knowledge, no study has reported prosthetic valve type as a potential factor for $^{18}$F-FDG PET/CT interpretation in PVE diagnosis. This feature could be implemented in the ML algorithm to evaluate its potential help in future work.

**STUDY LIMITATIONS.** The exclusive restriction to prosthetic aortic valves certainly represents a significant limitation of our work. The gold standard, on the basis of expert consensus among the endocarditis team 3 months after patient admission, does not allow reliable localization of the infectious focus. For patients with a single prosthetic valve, it is commonly acknowledged that the prosthetic valve is at greater risk of infection than the other valves, but for patients with 2 or more prosthetic valves, there is no certainty about location of the infection. Consequently, after assuming that patients with a single aortic prosthetic valve were the most numerous, we decided to include these patients exclusively to obtain the most accurate algorithm possible. As a result, our data cannot currently be extrapolated to patients with numerous prosthetic valves or a single prosthetic valve other than aortic. As mentioned by the Endopet study and American College of Cardiology/American Heart Association guidelines, visual interpretation of $^{18}$F-FDG PET/CT scans requires highly trained centers with high expertise in nuclear medicine.[2,6] Our ML algorithm required only a manual drawing of an ROI encompassing the prosthetic aortic valve, but this operation was performed retrospectively and in a single center with high cardiologic imaging experience, thus potentially limiting its use by less experienced centers. Furthermore, logistic regression, a classical classification model, was chosen to preserve the interpretability and to provide more insights when analyzing the results, but it requires precise image segmentation. Despite all of our results suggesting the relative reliability of our ML algorithm, this manual segmentation may have directly influenced both the construction and the final results of the algorithm.

## CONCLUSIONS

The ML algorithm reached diagnostic performance results close to those previously published that were based on visual assessment by highly trained experts, particularly when included in the ESC 2015 criteria. These results, obtained regardless of myocardial uptake, and the ease of use of our ML algorithm hold

up the prospect of a wider use of quantitative analysis of $^{18}$F-FDG PET/CT. If performance can be improved further, this approach seems very promising to optimize the role of $^{18}$F-FDG PET/CT in PVE diagnosis.

## FUNDING SUPPORT AND AUTHOR DISCLOSURES

This work was partially funded by the European Regional Development Fund, the Pays de la Loire region on the Connect Talent MILCOM programme (Multi-modal Imaging and Learning for Computational-based Medicine), Nantes Métropole (Convention 2017-10470), the French National Research Agency Investissements d'Avenir LabEx Iron (number ANR-11-LABX-0018-01), the Integrated Cancer Research Site (SIRIC) Imaging and Longitudinal Investigations to Ameliorate Decision Making (ILIAD) (INCa-DGOS-Inserm-12558), and the Clinical Research Institute TransForMed (I-SITE NExT). The authors have reported that they have no relationships relevant to the contents of this paper to disclose.

**ADDRESS FOR CORRESPONDENCE:** Dr Thomas Eugène, Department of Nuclear Medicine, CHU de Nantes, 1 Place Alexis Ricordeau, 44093 Nantes, France. E-mail: thomas.eugene@chu-nantes.fr.

## PERSPECTIVES

**COMPETENCY IN PATIENT CARE AND PROCEDURAL SKILLS:** $^{18}$F-FDG PET/CT quantitative analysis is a potentially easy to use aid for the diagnosis of PVE, but improvements are required to increase its reproducibility among centers. The information provided by the quantitative analysis were more powerful when integrated into a composite score combining clinical, biologic, and imaging information and therefore should be considered by the endocarditis team as an additional tool in the diagnosis of PVE.

**TRANSLATIONAL OUTLOOK:** Future prospective studies including deep learning approaches on larger groups will be needed to extend these promising results to other valves and further improve performance to provide a reliable and accurate diagnostic tool.

## REFERENCES

**1.** Habib G, Lancellotti P, Antunes MJ, et al. 2015 ESC guidelines for the management of infective endocarditis: the Task Force for the Management of Infective Endocarditis of the European Society of Cardiology (ESC). Endorsed by: European Association for Cardio-Thoracic Surgery (EACTS), the European Association of Nuclear Medicine (EANM). *Eur Heart J*. 2015;36:3075–3128.

**2.** Otto C, Nishimura R, Bonow R, et al. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: executive summary. *J Am Coll Cardiol*. 2021;77:450–500.

**3.** Saby L, Laas O, Habib G, et al. Positron emission tomography/computed tomography for diagnosis of prosthetic valve endocarditis: increased valvular $^{18}$F-fluorodeoxyglucose uptake as a novel major criterion. *J Am Coll Cardiol Img*. 2013;61:2374–2382.

**4.** Pizzi MN, Roque A, Fernández-Hidalgo N, et al. Improving the diagnosis of infective endocarditis in prosthetic valves and intracardiac devices with $^{18}$F-fluordeoxyglucose positron emission tomography/computed tomography angiography: initial results at an infective endocarditis referral center. *Circulation*. 2015;132:1113–1126.

**5.** Anton-Vazquez V, Cannata A, Amin-Youssef G, et al. Diagnostic value of $^{18}$F-FDG PET/CT in infective endocarditis. *Clin Res Cardiol*. 2022;111:673–679.

**6.** Philip M, Tessonier L, Mancini J, et al. Comparison between ESC and Duke criteria for the diagnosis of prosthetic valve infective endocarditis. *J Am Coll Cardiol Img*. 2020;13:2605–2615.

**7.** Swart LE, Gomes A, Scholtens AM, et al. Improving the diagnostic performance of $^{18}$F-fluorodeoxyglucose positron-emission tomography/computed tomography in prosthetic heart valve endocarditis. *Circulation*. 2018;138:1412–1427.

**8.** Roque A, Pizzi MN, Fernández-Hidalgo N, et al. The valve uptake index: improving assessment of prosthetic valve endocarditis and updating [$^{18}$F] FDG PET/CT(A) imaging criteria. *Eur Heart J Cardiovasc Imaging*. 2022;23:1260–1271.

**9.** Scholtens AM, Swart LE, Kolste HJT, Budde RPJ, Lam MGEH, Verberne HJ. Standardized uptake values in FDG PET/CT for prosthetic heart valve endocarditis: a call for standardization. *J Nucl Cardiol*. 2018;25:2084–2091.

**10.** Mannil M, Eberhard M, von Spiczak J, Heindel W, Alkadhi H, Baessler B. Artificial intelligence and texture analysis in cardiac imaging. *Curr Cardiol Rep*. 2020;22:131.

**11.** van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77:e104–e107.

**12.** Bailly C, Bodet-Milin C, Couespel S, et al. Revisiting the robustness of PET-based textural features in the context of multi-centric trials. *PLoS One*. 2016;11:e0159984.

**13.** Desseroit MC, Tixier F, Weber WA, et al. Reliability of PET/CT shape and heterogeneity features in functional and morphologic components of non-small cell lung cancer tumors: a repeatability analysis in a prospective multicenter cohort. *J Nucl Med*. 2017;58:406–411.

**14.** van Velden FHP, Kramer GM, Frings V, et al. Repeatability of radiomic features in non-small-cell lung cancer [(18)F]FDG-PET/CT studies: impact of reconstruction and delineation. *Mol Imaging Biol*. 2016;18:788–795.

**15.** Pfaehler E, Mesotten L, Zhovannik I, et al. Plausibility and redundancy analysis to select FDG-PET textural features in non-small cell lung cancer. *Med Phys*. 2021;48:1226–1238.

**16.** Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority oversampling technique. *J Artif Intell Res*. 2002;16:321–357.

**17.** He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng*. 2009;21:1263–1284.

**18.** Sengupta PP, Shrestha S, Berthon B, et al. Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME): a checklist: reviewed by the American College of Cardiology Healthcare Innovation Council. *J Am Coll Cardiol Img*. 2020;13:2017–2035.

**19.** Gomes A, Glaudemans AWJM, Touw DJ, et al. Diagnostic value of imaging in infective endocarditis: a systematic review. *Lancet Infect Dis*. 2017;17:e1–e14.

**20.** Cecchi E, Parrini I, Chinaglia A, et al. New diagnostic criteria for infective endocarditis. A study of sensitivity and specificity. *Eur Heart J*. 1997;18:1149–1156.

**21.** Dilsizian V, Budde RPJ, Chen W, Mankad SV, Lindner JR, Nieman K. Best practices for imaging cardiac device-related infections and endocarditis: a *JACC: Cardiovascular Imaging* expert panel statement. *J Am Coll Cardiol Img*. 2022;15:891–911.

**22.** Boursier C, Duval X, Mahida B, et al. Hypermetabolism of the spleen or bone marrow is an additional albeit indirect sign of infective endocarditis at FDG-PET. *J Nucl Cardiol*. 2021;28:2533–2542.

**23.** Scholtens AM, van Aarnhem EEHL, Budde RP. Effect of antibiotics on FDG-PET/CT imaging of prosthetic heart valve endocarditis. *Eur Heart J Cardiovasc Imaging*. 2015;16:1223.

**KEY WORDS** artificial intelligence, endocarditis, FDG, PET

**APPENDIX** For expanded Methods and Results sections as well as supplemental figures, tables, and references, please see the online version of this paper.