EJNMMI Reports

ORIGINAL ARTICLE Open Access

Practical use of radiomic features as a metric for image quality discrimination in [^{18}F] FDG-PET: a pilot study

Jane Burns[1*], Hannah O'Driscoll[2] and Eamon Loughman[2]

*Correspondence:
jane.burns@umail.ucc.ie

[1] Department of Radiology, Mater Misericordiae University Hospital, Eccles St, Phibsborough, Dublin 7 D07 R2WY, Ireland
[2] Department of Medical Physics, Mater Private Network, Eccles St, Phibsborough, Dublin, Ireland

Abstract

Purpose: Radiomics features have been utilised as group metrics of image quality in many areas of diagnostic radiology. In this pilot study, the relationship between technical metrics used in image quality assurance and visual grading scores provided by a radiologist were evaluated. Image dataset harmonisation allowed comparison between the two and allowed trends to be extracted. We propose a reproducible technique to identify the metrics.

Methods: A retrospective chart review of 30 [^{18}F] FDG-PET/CT performed in a nuclear medicine referral centre was performed. Image datasets were reprocessed to correspond to a bed duration of 180, 120, 60, 30 s per bed position and were analysed according to a pre-set bank of semi-quantitative features by a radiology resident. The extraction of radiomic features in PET images was performed using SLICER-RADIOMICS Module version 5.2.2. To facilitate the comparison of radiomic features and radiologist scoring data, normalisation was performed on both data sets. Fréchet distance analysis, Mean Square Error and Mean Absolute Error display the level of agreement between features and radiologist following the rescale of the data.

Results: Of the 120 reprocessed image datasets, 115 were included in the study. We focused on overall image quality score rather than individual radiomic metrics as this identified the most robust trend. A significant difference in the 30 s image dataset with respect to each group individually and combined for the radiologist overall score was observed.

Conclusion: Our results show that a large percentage change in certain features can indicate a significant change in quality in clinically processed images.

Keywords: Radiomics, Image quality, [^{18}F] FDG-PET/CT, Quantitative imaging

Purpose

Positron emission tomography (PET) has been in use clinically for since the 1970s (López-Mora et al. 2022). In recent years, digital PET has been developed and is used clinically. However, there are several drawbacks to the technique, such a poor spatial resolution, noise due to scattered photons and reduced signal, as a result of systems dead-time. A low signal to noise ratio (SNR) is a consequence of these (Singh 2024; Gong

et al. 2021). The introduction of silicon photomultipliers (SiPMs) to replace photomultiplier tubes (PMTs), the spatial and coincidence time resolutions have increased due to increased detector sensitivity (Fragoso Costa et al. 2022). SiPMs achieve a 1-to-1 correspondence between photomultipliers and detectors. This results in a very good intrinsic timing resolution and increased dynamic range, meaning that SiPMs improve both the sensitivity and the maximum count rate. Thus, improved image resolution.

There is a growing trend towards quantitative imaging, with increased optimism that identifying and quantifying image-derived biomarkers will enhance the capabilities of PET and lead to better patient outcomes. For example, many studies (Cook et al. 2013; Tixier et al. 2011) on quantitative PET have yielded promising outcomes in tumour texture analysis, suggesting that radiomic features can effectively capture tumour heterogeneity and potentially be used to predict clinical outcomes. Radiomics has been utilised in image quality in many areas of diagnostic radiology (Reynés-Llompart et al. 2019; Whybra, et al. 2024; Amakusa et al. 2018; Mizuta et al. 2009). While the term is not strictly defined (Mayerhoefer et al. 2020), Radiomics is a method that aims to extract large-set quantitative image-derived metrics, such as intensity and texture, from diagnostic images, that are ideally reproducible, to identify associations between imaging data and patient outcomes. However, clinical application and adoption has been limited, which may be partly due to overall poor generalisability and lack of robustness. Many radiomic metrics are sensitive to different acquisition protocols (Galavis et al. 2010), different scanners (Reuzé et al. 2017) and reconstruction algorithms (Yan et al. 2015). Unfortunately, this reflects real-world clinical practice, where there is diversity in image acquisition across clinical sites and variables like acquisition protocols and scanners cannot be completely standardised. Nonetheless, they are useful tools for image quality assessment in [^{18}F] FDG-PET/CT and nuclear medicine.

Radiographic image quality may be considered as a balance between three primary parameters: noise, spatial resolution and contrast (Mayo et al. 2003; Mattsson and Söderberg 2011. In [^{18}F] FDG-PET/CT, it is a function of both dose and acquisition time. Image quality is an important factor in interpreting and accurately identifying pathology. Noise is unavoidable in medical imaging and is often a major factor limiting image quality (McNitt-Gray 2006; Primak et al. 2006). An increase in noise reduces both low-contrast and spatial resolution, affecting how radiologists perceive the image. However, a certain level of noise is tolerable within a system and may not significantly impair image interpretation (Kalender 2011; Wessling et al. 2003). The standard of image quality is multifactorial, considers radiologists' preferences, as well as striking a balance between acceptable image noise and radiation dose.

Observer performance methods, including visual grading analysis (VGA), are established techniques for assessing and analyzing image quality in medical imaging and diagnostic radiology (Ludewig et al. 2010). The validity of VGA is considered high, as observer ratings take into consideration experience and confidence interpreting an image and technical factors. Image assessment is conducted by evaluating the visibility of clinically relevant structures, following established standards like the European guidelines on quality criteria. Reliable validated methods for assessing image quality are essential for the implementation of optimised PET protocols. VGA by radiologists may be useful in identifying simple, reproducible radiomic features (Piñeiro-Fiel et al. 2021)

which may act as quick image quality control assessment, prior to nuclear medicine specialist identification of a suboptimal study. Simple, freely available metrics which are quickly calculated could potentially increase the likelihood of widespread adoption and minimise additional expense to users.

In this study, the relationship between technical metrics used in image quality assurance and visual grading scores provided by a radiologist were evaluated. Our objective in this pilot study is to identify reproducible radiomics features that have been scaled to radiologist VGA, as potential image quality discriminators in [^{18}F] FDG-PET. VGA was performed by a resident radiologist on EANM Research Ltd (EARL)-accredited images and radiomics assessment was performed by medical physicists. Traditionally harmonisation has been performed between sites, the harmonisation in this study was between VGA and technical radiomic metrics. Image dataset harmonisation allowed comparison between the two allowed trends to be extracted. We propose a reproducible technique to identify the metrics.

Methods

Patient population

This is a pilot retrospective chart review of 30 patients referred for [^{18}F] FDG-PET/CT to the Mater Private Hospital Network was conducted between August to September 2022. The male-to-female ratio was 1:2. There was a broad range of patient weights between 48 and 112.5 kg (kg), with an average weight of 74 kg. The range of injected activity was between 108 and 273 megabecquerel (MBq), with a mean injected activity of 187 MBq. Patient diabetes mellitus status was recorded and a cut-off of acceptable blood glucose level of 6 was used. Our inclusion criteria was broad, as a representative sample of our patient cohort was desired. Patients with a variety of clinical indications for [^{18}F] FDG-PET were included, for example, assessment and staging of various malignancies. Studies that included whole-body, from mid-orbital to mid-thigh, [^{18}F] FDG-PET images that were scanned for at least 3 min per bed in a list mode technique with intact raw data for reprocessing purposes. Any deviations from normal departmental whole body [^{18}F] FDG-PET/CT scanning protocol were excluded. Dedicated brain [18F] FDG-PET studies, incomplete studies, studies that were scanned for less than 3 min per bed and those with corrupted raw data were excluded.

Image acquisition

Whole body image (Mid Orbital to Mid Thigh) List-mode data was acquired using a Vision 450 PET/CT scanner (Siemens Healthineers) with a bed duration of 3 min per position. In accordance with normal departmental guidelines, scans were acquired 60 min after injection of [^{18}F] FDG, administration method was weight based with a guideline of 2.5 MBq/kg. Attenuation-,Random- and scatter-corrected PET data were reconstructed iteratively including point-spread function and time-of-flight information using the TrueX algorithm (Siemens Medical Solutions); the settings were 4 iterations, 5 subsets, 4.5 mm gaussian filter. A diagnostic CT scan was obtained before the PET scan with a 3.0 mm (16 X 1.2 mm) slice thickness, 120 kV, and a quality reference of 160 mAs modulated by the Care Dose automatic exposure control system.

Image analysis

Initial image quality assessments were performed using the NEMA IEC phantom. Radiomics was completed on the images by analysing the texture of specific regions. For the NEMA IEC image quality phantom, regions of 30 mm in diameter were chosen on the large sphere, lung insert and background regions. This diameter was due to the diameter of the large sphere being 37 mm with some room for error to avoid the surrounding background region. For all further radiomics analysis all ROIs were 30 mm in diameter.

Thirty patient image data sets were reprocessed to create datasets of 180, 120, 60, 30 s per bed position. This resulted in a total of 120 datasets for analysis, four datasets of 30 reprocessed [^{18}F] FDG-PET images for corresponding bedtime. Images were first deidentified by removal of all patient identifying Digital Imaging and Communications in Medicine (DICOM) headers and assigned a non-consecutive serial number as an identifier. Deidentified and reprocessed images were analysed according to a pre-set bank of semi-quantitative features by a radiology resident. The images were graded on image quality on a five-point scales in various categories. Firstly, an overall score indicating if the image quality is acceptable for diagnostic purposes. The quality and ability to interpret the attenuation correction (AC) acquisition was also evaluated. The ability to identify and interpret for lymph node involvement, skull base and neck, thorax and abdomen, and pelvis to mid-thigh. Additionally the standardised uptake value (SUV) of the liver was recorded.

The extraction of radiomic features in PET images was performed using SLICER-RADIOMICS Module version 5.2.2. A segmentation of 30 mm in diameter (bin width 25) was added to the central liver slice and in each neighbouring axial slice. The average of these three segmentations were found. Using these values an average value was found for each scan time to generate the radiomics metrics.

Statistical analysis

To facilitate the comparison of radiomic features and radiologist scoring data normalisation was performed on both data sets. Fréchet distance analysis, Mean Square Error and Mean Absolute Error display the level of agreement between features and radiologist following the rescale of the data. This type of rescaling of features has good evidence and implementation in the field of radiomics as found in similar implementation in ComBat methodology (Horng, et al. 2022). Features which had good agreement according to the above metrics were then analysed using single factor ANOVA analysis for significance, planned post hoc t-test intercomparison of specific group interactions were accounted for using the Bonferroni correction method to adjust the level of significance acceptable. A P value of less than 0.0125 was considered statistically significant. All analyses were performed using Microsoft Excel [365 v.2402] and Python libraries numpy, matplotlib.pyplot, pandas, sklearn, scipy (Python version 3.8.3).

Results

Of the 120 reprocessed image datasets, 115 (95.8%) were included in the study. 5 image datasets were corrupted during reprocessing and accurate SUV and features could not be included in the assessment. We focused on overall image quality score rather than

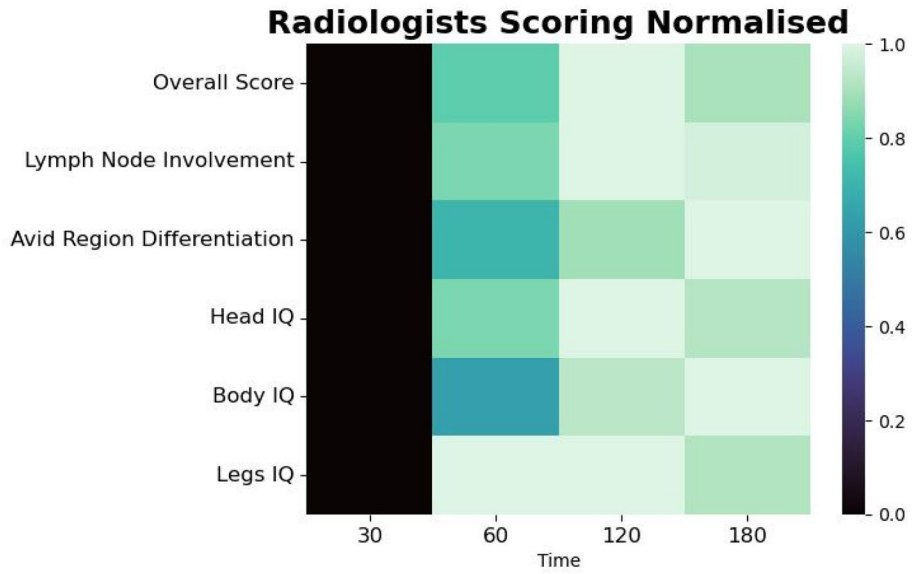

Fig. 1 Normalised radiologist scoring for each time dataset

Table 1 Cluster Prominence

Metric	30 V all significance	30 V 1 min significance	30 V 2 min significance	30 V 3 min significance
ANOVA SINGLE FACTOR (significance < 0.05)				
Cluster prominence	1.3163×10^{-6}	0.0121	0.0024	0.0018
Cluster shade	0.0001	0.0639	0.0151	0.0104
Kurtosis	0.0426	0.1640	0.0514	0.1425
Complexity	4.4316×10^{-15}	9.0339×10^{-5}	9.7514×10^{-9}	2.8636×10^{-9}
Autocorrelation	4.2812×10^{-17}	6.8124×10^{-6}	1.3314×10^{-10}	4.4922×10^{-12}
High gray level Emphasis	4.7262×10^{-17}	7.8345×10^{-6}	1.4113×10^{-10}	4.8176×10^{-12}
T-TESTS (significance < 0.0125)				
Cluster prominence	1.3163×10^{-6}	0.0121	0.0024	0.0018
Cluster shade	0.0001	0.0639	0.0151	0.0104
Kurtosis	0.0426	0.1640	0.0514	0.1425
Complexity	4.4316×10^{-15}	9.0339×10^{-5}	9.7514×10^{-9}	2.8636×10^{-9}
Autocorrelation	4.2812×10^{-17}	6.8124×10^{-6}	1.3314×10^{-10}	4.4922×10^{-12}
High gray level Emphasis	4.7262×10^{-17}	7.8345×10^{-6}	1.4113×10^{-10}	4.8176×10^{-12}

individual radiomic metrics as this identified the most robust trend. Below is a heat map of demonstrating the radiologist scoring (Fig. 1). There is a noticeable decline in image quality and ability to interpret relevant structures at the thirty second bed time.

Radiologist grading

No significant differences were found between 60, 120 and 180 s image datasets, with overall score and ability to interpret various regions of interest being relatively similar (Table 1). The overall images quality and diagnostic interpretation was acceptable.

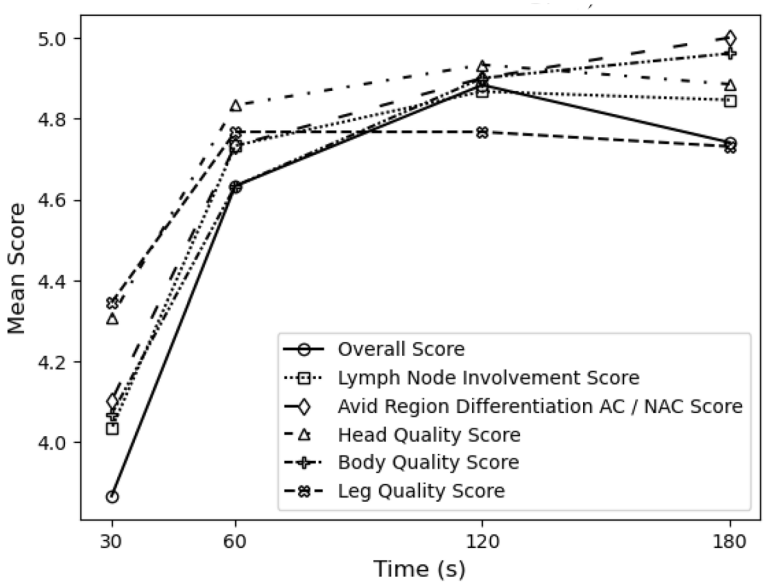

Fig. 2 Mean radiologist semi-quantitative metrics

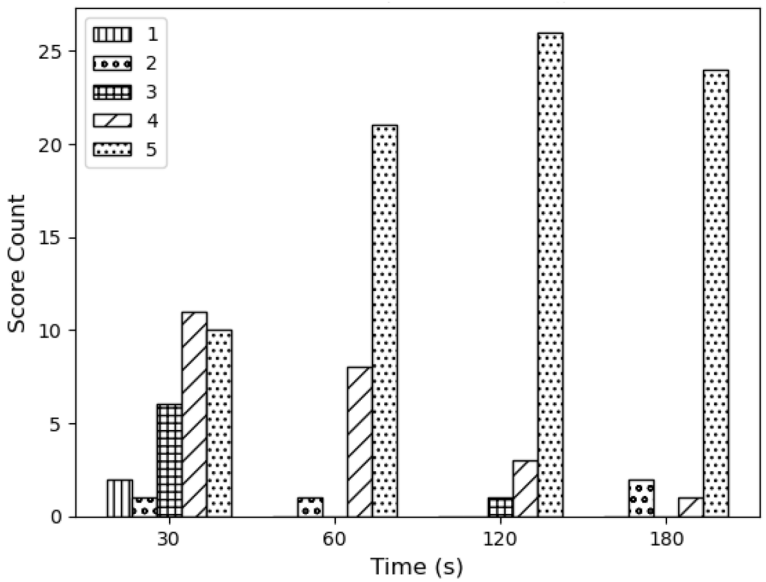

Fig. 3 Radiologist score count for each bed time

However, a significant difference in the 30 s image dataset with respect to each group individually and combined for the radiologist overall score is demonstrated (Fig. 2). In addition, metric variability was found to be far more pronounced in the 30 s image datasets (Fig. 3).

For bed times between 60 and 180 s, there was no significance reported for the overall image quality, lymph node involvement score or head and legs image quality scores (Fig. 4). The scores given to the image quality of the body has the most significance associated with it. Using further t-tests, there was no significance between the grading for

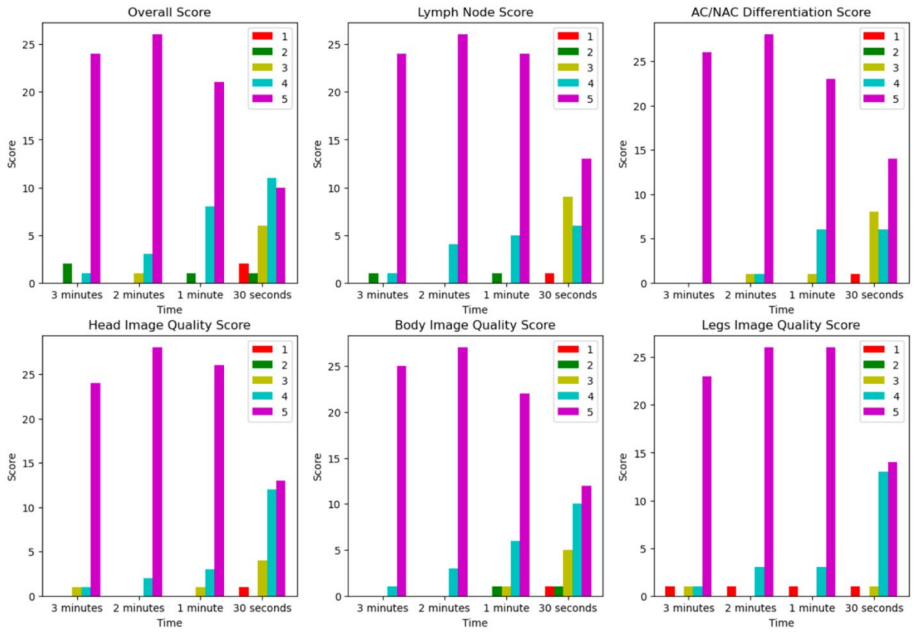

Fig. 4 Individual metrics radiologist grading

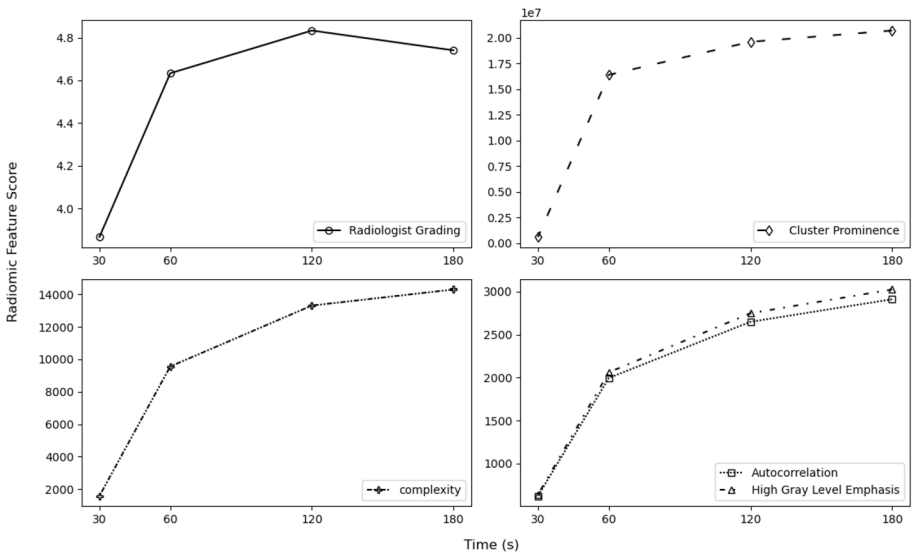

Fig. 5 Best radiomic features and radiologist grading on original scale

180 and 120 s bed times or the grading for 120 and 60 s. However there still was a significant difference between the 60 s and 30 s bed times.

Radiomic analysis

Signal similarity metrics Fréchet distance and Mean Square Error (MSE) were used to match radiomic feature performance trend to radiologist scoring trend (Fig. 5). These metrics were combined and demonstrate a marked reduction in acceptability at 30 s (Fig. 6). Curve inversion was applied on all relevant feature datasets to assess and issues

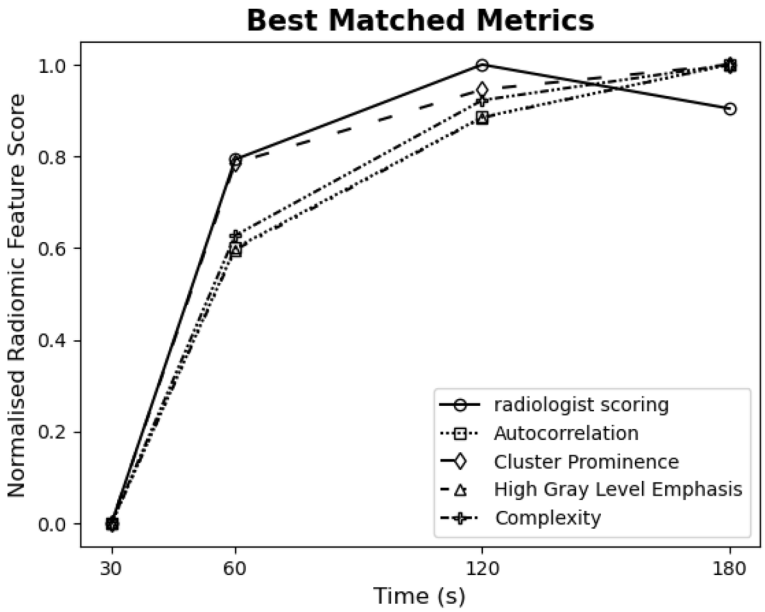

Fig. 6 Best matched metrics

with orientation or directionality (Fig. 7a, b). A Fréchet distance cutoff was set at 0.3, while a corresponding MSE cutoff was set at 0.1.

Significant differences were found between several radiomic features in the clinical image datasets. Both scaled and unscaled data for the radiomic features which satisfied both similarity and significance requirements were found to show large differences between 30 s dataset and all others (Table 2). Additionally, there is an overall downward trend in various radiomic features as bedtimes decrease, most notably at 30 s (Table 3).

Discussion

In this pilot study, we focused only on [^{18}F] FDG-PET EARL standard 2 images to assess image quality. The reprocessed patient datasets allowed both semi quantitative radiologist analysis and quantitative radiomic feature analysis. Harmonisation of the results of the VGA and radiomics features was performed to allow for comparison between the two. However, as all images were performed on the same scanner with the same standard departmental acquisition protocol to yield EARL standard 2 images, harmonisation of the resultant images was not required.

Quantitative PET imaging has emerged as a promising approach that offers additional insights into prognosis and treatment response, surpassing the capabilities of traditional qualitative imaging. However, the sensitivity of quantitative metrics to different scanners, acquisition protocols, and reconstruction algorithms is a limiting factor. Going forward, it would be pertinent to confirm our findings across multiple sites. In this case, harmonisation of both VGA and image processing would need to be performed between sites. However, we propose that EARL-accredited centres could use our techniques and produce similar outcomes.

Trends in radiomic metrics and VGA were compared using signal similarity metrics and ANOVA analysis. Significant differences in image quality identified by radiologist

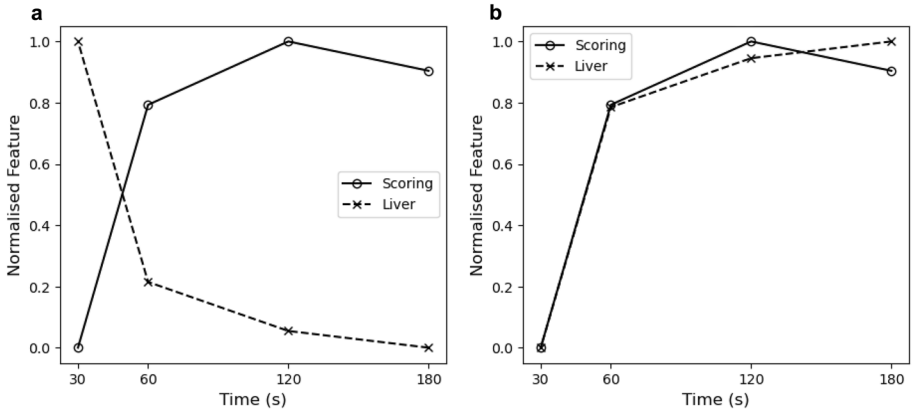

Fig. 7 **a** Original Curve: Cluster Prominence and **b** Inverted Curve: Cluster Prominence

Table 2 Prominence

Metric	30 s	1 min	2 min	3 min
SCALED				
Cluster prominence	0	0.785	0.945	1
Complexity	0	0.627	0.922	1
Autocorrelation	0	0.600	0.885	1
High gray level Emphasis	0	0.596	0.884	1
UNSCALED				
Cluster prominence	6.083×10^5	1.639×10^7	1.961×10^7	2.070×10^7
Complexity	1584	9559	13,310	14,299
Autocorrelation	623	1995	2648	2911
High gray level Emphasis	640	2061	2748	3025

Table 3 Feature

Radiomic feature	% Decrease 180–30 s	% Decrease 120–30 s	% Decrease 60–30 s
Cluster prominence	97.06	96.9	96.29
Complexity	88.92	88.1	83.43
Autocorrelation	78.59	76.47	68.77
High grey level Emphasis	78.84	76.71	68.94

assessment were mirrored in radiomic analysis of specific features, in particular the overall image score. There was no significant difference between the average scoring of image quality at the 180, 120 and 60 s bed times. However, we have suggested that there is a significant decline in image quality and hence clinical interpretability when bed times are reduced to 30 s. The liver region is used as a background region due to its uniform uptake, as well as being a quick and relative easy reference point. This makes it a useful metric to normalise uptake in other regions of interest throughout the body.

Although the scope of this study is limited due to the small sample size, this was intended to be a pilot study to test our hypothesis. The VGA was performed by one radiology resident alone and was not asked to comment or evaluate specific pathologies that may be present. Going forward, increasing the number of radiologists to review images would be required. The inherent low spatial resolution of [^{18}F] FDG-PET can itself lead to difficulties with analysis, especially textural analysis when compared with other imagine modalities.

The identification of favourable radiomic metrics using our proposed technique, provides valuable insights for future research and clinical applications. These metrics serve as objective measures for assessing image quality and guiding optimization efforts in [^{18}F] FDG-PET/CT imaging protocols. The proposed binary metric percentage reduction method offers a practical tool to classify images as acceptable or unacceptable based on predefined cut-off points. This highlights the potential of radiomics in streamlining image interpretation processes and enhancing diagnostic accuracy in clinical practice.

Conclusion

In this study we explore the identification and use of individual radiomic features as binary metrics of image quality. Our results show that a large percentage change in certain features can indicate a significant change in quality in clinically processed images.

Acknowledgements
The authors would like to acknowledge Prof. Martin O'Connell, consultant radiologist and nuclear medicine specialist for his support.

Author contributions
J.B. was the radiology resident who reviewed and scored the imaging, wrote the manuscript and editing. H.OD. processed the data and results. E.L. acted as supervising author, reviewing the data and the manuscript. All authors read and approved the final manuscript.

Funding
The authors did not receive funding from any organization for the submitted work.

Availability of data and material
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Mater Hospital Private Network Ethics Committee reviewed this study and waived the need for ethics approval.

Consent for publication
Images used during the study were entirely unidentifiable and there are no details on individuals reported within the manuscript. Patient consent was sought for use of their imaging for the study and for reconstructions for review.

Competing interests
The authors declare that they have no competing interests.

Received: 18 June 2024 Accepted: 29 January 2025
Published online: 08 May 2025

References

Amakusa S, Matsuoka K, Kawano M et al (2018) Influence of region-of-interest determination on measurement of signal-to-noise ratio in liver on PET images. Ann Nucl Med 32:1–6. https://doi.org/10.1007/s12149-017-1215-y
Cook GJ, Yip C, Siddique M, Goh V, Chicklore S, Roy A, Marsden P, Ahmad S, Landau D (2013) Are pretreatment 18F-FDG PET tumor textural features in non–small cell lung cancer associated with response and survival after chemoradiotherapy? J Nucl Med 54(1):19–26. https://doi.org/10.2967/jnumed.112.107375

Fragoso Costa P, Jentzen W, Brahmer A, Mavroeidi IA, Zarrad F, Umutlu L, Fendler WP, Rischpler C, Herrmann K, Conti M, Seifert R (2022) Phantom-based acquisition time and image reconstruction parameter optimisation for oncologic FDG PET/CT examinations using a digital system. BMC Cancer 22(1):899. https://doi.org/10.1186/s12885-022-09993-4

Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R (2010) Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. Acta Oncol 49(7):1012–1016. https://doi.org/10.3109/0284186X.2010.498437

Gong K, Kim K, Cui J, Wu D, Li Q (2021) The evolution of image reconstruction in PET: from filtered back-projection to artificial intelligence. PET Clinics 16(4):533–542. https://doi.org/10.1016/j.cpet.2021.06.004

Horng H et al (2022) Generalized ComBat harmonization methods for radiomic features with multi-modal distributions and multiple batch effects. Sci Rep. https://doi.org/10.1038/s41598-022-08412-9

Kalender WA (2011) Computed tomography: fundamentals, system technology, image quality, applications. John Wiley & Sons

López-Mora DA, Carrió I, Flotats A (2022) Digital PET vs analog PET: clinical implications? Semin Nucl Med. https://doi.org/10.1053/j.semnuclmed.2021.10.004

Ludewig E, Richter A, Frame M (2010) Diagnostic imaging–evaluating image quality using visual grading characteristic (VGC) analysis. Vet Res Commun 34:473–479. https://doi.org/10.1007/s11259-010-9413-2

Mattsson S, Söderberg M (2011) Radiation dose management in CT, SPECT/CT and PET/CT techniques. Radiat Prot Dosim 147(1–2):13–21. https://doi.org/10.1093/rpd/ncr261

Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, Cook G (2020) Introduction to radiomics. J Nucl Med 61(4):488–495. https://doi.org/10.2967/jnumed.118.222893

Mayo JR, Aldrich J, Müller NL (2003) Radiation exposure at chest CT: a statement of the Fleischner society. Radiology 228(1):15–21. https://doi.org/10.1148/radiol.2281020874

McNitt-Gray MF (2006) Tradeoffs in CT image quality and dose. Med Phys 33(6):2154–2162. https://doi.org/10.1118/1.2241390

Mizuta T, Senda M, Okamura T et al (2009) NEC density and liver ROI S/N Ratio for image quality control of whole-body FDG-PET scans: comparison with visual assessment. Mol Imaging Biol 11:480–486. https://doi.org/10.1007/s11307-009-0214-3

Piñeiro-Fiel M, Moscoso A, Pubul V, Ruibal Á, Silva-Rodríguez J, Aguiar P (2021) A systematic review of PET textural analysis and radiomics in cancer. Diagnostics 11(2):380. https://doi.org/10.3390/diagnostics11020380

Primak AN, McCollough CH, Bruesewitz MR, Zhang J, Fletcher JG (2006) Relationship between noise, dose, and pitch in cardiac multi–detector row CT. Radiographics 26(6):1785–1794. https://doi.org/10.1148/rg.266065063

Reuzé S, Orlhac F, Chargari C, Nioche C, Limkin E, Riet F, Escande A, Haie-Meder C, Dercle L, Gouy S, Buvat I (2017) Prediction of cervical cancer recurrence using textural features extracted from 18F-FDG PET images acquired with different scanners. Oncotarget 8(26):43169. https://doi.org/10.18632/oncotarget.17856

Reynés-Llompart G, Sabaté-Llobera A, Llinares-Tello E et al (2019) Image quality evaluation in a modern PET system: impact of new reconstructions methods and a radiomics approach. Sci Rep 9:10640. https://doi.org/10.1038/s41598-019-46937-8

Singh MK (2024) A review of digital PET-CT technology: comparing performance parameters in SiPM integrated digital PET-CT systems. Radiography 30(1):13–20. https://doi.org/10.1016/j.radi.2023.10.004

Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, Corcos L, Visvikis D (2011) Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med 52(3):369–378. https://doi.org/10.2967/jnumed.110.082404

Wessling J, Fischbach R, Meier N, Allkemper T, Klusmeier J, Ludwig K, Heindel W (2003) CT colonography: protocol optimization with multi–detector row CT—study in an anthropomorphic colon phantom. Radiology 228(3):753–759. https://doi.org/10.1148/radiol.2283020928

Whybra P et al (2024) The image biomarker standardization initiative: Standardized convolutional filters for reproducible radiomics and enhanced clinical insights. Radiology. https://doi.org/10.1148/radiol.231319

Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, Tham IW, Townsend D (2015) Impact of image reconstruction settings on texture features in 18F-FDG PET. J Nucl Med 56(11):1667–1673. https://doi.org/10.2967/jnumed.115.156927

Publisher's Note