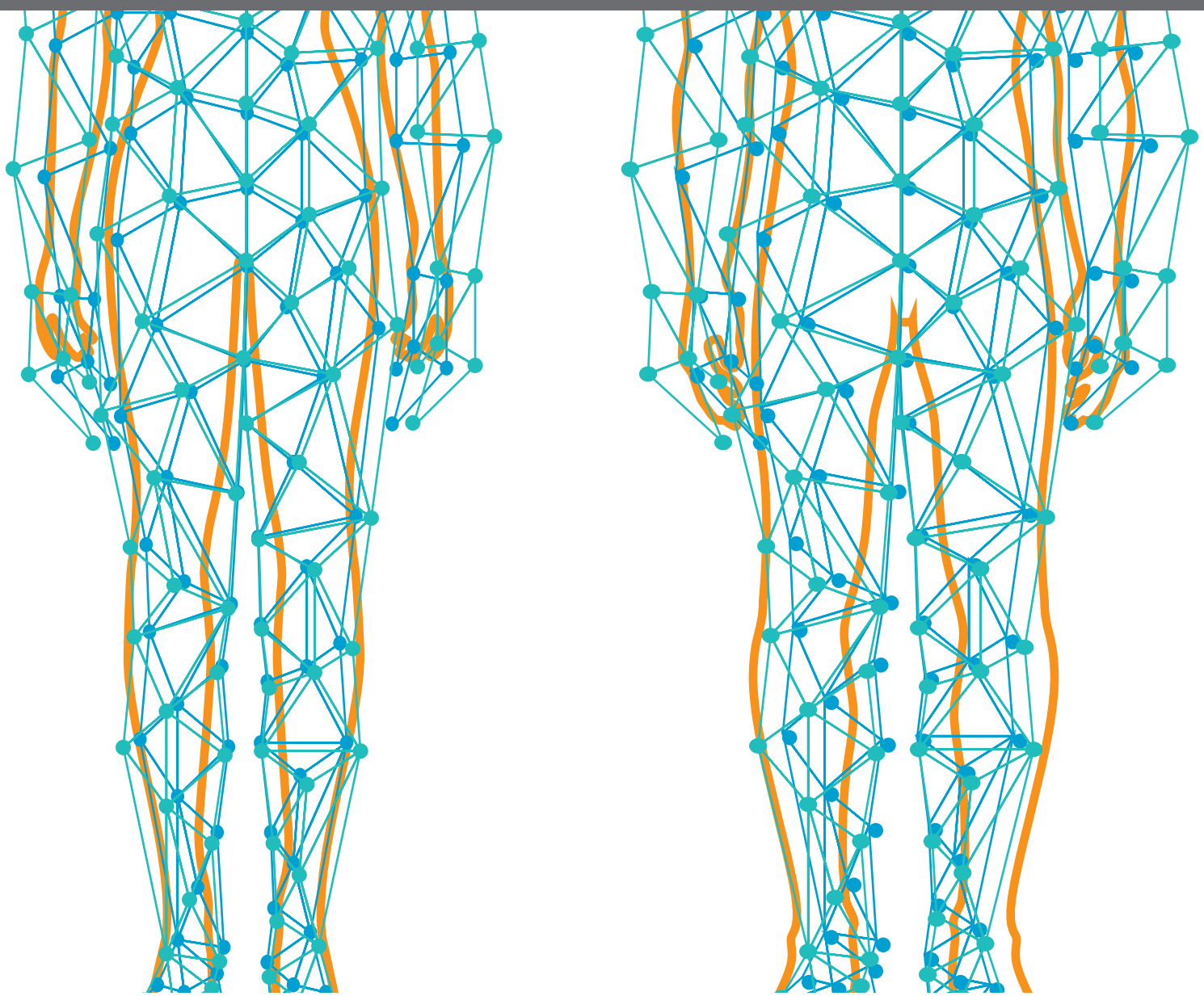
ARTIFICIAL INTELLIGENCE IN POSITRON EMISSION TOMOGRAPHY

EDITED BY: Xiaoli Lan, Kuangyu Shi, Xiu Ying Wang and Chuantao Zuo
PUBLISHED IN: Frontiers in Medicine



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-553-1
DOI 10.3389/978-2-88974-553-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ARTIFICIAL INTELLIGENCE IN POSITRON EMISSION TOMOGRAPHY

Topic Editors:
Xiaoli Lan, Huazhong University of Science and Technology, China
Kuangyu Shi, University of Bern, Switzerland
Xiu Ying Wang, The University of Sydney, Australia
Chuantao Zuo, Fudan University, China

Citation: Lan, X., Shi, K., Wang, X. Y., Zuo, C., eds. (2022). Artificial Intelligence in Positron Emission Tomography. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-553-1

Table of Contents

frontiers in Medicine

EDITORIAL
published: 31 January 2022
doi: 10.3389/fmed.2022.848336

Editorial: Artificial Intelligence in Positron Emission Tomography

Hanyi Fang[1,2], *Kuangyu Shi*[3,4], *Xiuying Wang*[5], *Chuantao Zuo*[6] *and Xiaoli Lan*[1,2*]

[1] Department of Nuclear Medicine, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2] Hubei Province Key Laboratory of Molecular Imaging, Wuhan, China, [3] Department of Nuclear Medicine, University of Bern, Bern, Switzerland, [4] Department of Informatics, Technical University of Munich, Munich, Germany, [5] School of Computer Science, The University of Sydney, Sydney, NSW, Australia, [6] PET Center and National Clinical Research Center for Aging and Medicine, Huashan Hospital, Fudan University, Shanghai, China

Keywords: artificial intelligence, molecular imaging, positron emission tomography, oncology, neurology

Editorial on the Research Topic

Artificial Intelligence in Positron Emission Tomography

Smartphones, smart homes, and intelligent navigation are all examples of important applications of artificial intelligence (AI) in our daily life. AI was initially introduced in the 1950s, with the development of understanding and redefinition. AI is currently defined as a new technological science that studies and develops theorems, methods, technologies, and application systems that are used to simulate, extend, and enhance human intelligence (1).

We have witnessed the rapid advancement of AI, and its research and application in medical care, especially processing and analysizing the medical images is in the ascendant. In comparison to computed tomography (CT) and magnetic resonance imaging (MRI), which are more accessible and easier to standardize the acquisition processes, the positron emission tomography (PET) is more expensive and less broadly accessible, and its more complicated technical operation process poses difficulty on standardizing the image acquisition. Though the research and application of AI in PET is relatively slower, since PET is such an essential field of molecular imaging, AI in PET imaging is attracting substantial research attention and becoming a research hotspot. At the level of technology, image post-processing, including image standardization, normalization, wavelet transformation, Gaussian transformation, and feature preprocessing, have been studied with aims to solve the challenges posed by the parameter and quality variations and differences when imaging with PET scanners from different manufacturers, instrument models, and imaging technologies. The AI-empowered segmentation techniques have further improved the stability of AI features and the repeatability of AI researches (2, 3). To address the needs of clinical applications, by mining deeply into image features, combining population and clinical evidence, and constructing machine learning models, AI in PET has been developed for lesion detection and boundary delineation, diagnosis and differential diagnosis, risk prediction and prognostic evaluation, and even the prediction of clinical gene or molecular typing (1, 4–7).

This Research Topic comprises 11 publications that emphasized how AI supports PET image processing and analysis. Recently, numerous research groups have been focusing on the use of AI in PET image interpretation, such as lesion detection. Kawakami et al. applied an object deep learning (DL) detection model, You Only Look Once Version 2 (YOLOv2), to detect the physiological and abnormal uptake in ^{18}F-FDG PET. Results showed that the physiological uptake on MIP images was recognized quickly and precisely (Kawakami et al.). The abnormal uptake detected by YOLOv2 was with a high coverage rate of that manually identified (Kawakami et al.). The precise detection and fast response would be a useful tool in disease diagnosis. The maximal standardized uptake value (SUVmax) is the most commonly used parameter to interpret images and evaluate

OPEN ACCESS

Edited and reviewed by:
Giorgio Treglia,
Ente Ospedaliero Cantonale (EOC), Switzerland

***Correspondence:**
Xiaoli Lan
xiaoli_lan@hust.edu.cn

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: *04 January 2022*
Accepted: *07 January 2022*
Published: *31 January 2022*

Citation:
Fang H, Shi K, Wang X, Zuo C and Lan X (2022) Editorial: Artificial Intelligence in Positron Emission Tomography. Front. Med. 9:848336. doi: 10.3389/fmed.2022.848336

lesions in the daily diagnostic reports. In a preliminary study, Hirata et al. sought to define a precise SUVmax as an identifier to locate lesions. Although it is difficult to identify the lesions when the SUVmax < 2 in ^{18}F-FDG PET scans, this approach could help for the construction of the AI training dataset (Hirata et al.).

In oncology, AI has been used for diagnosis, differential diagnosis, and cancer staging. Satoh et al. used texture analysis in a retrospective study to compare the diagnosis ability of breast cancer between the high-resolution dedicated breast PET (dbPET) and whole-body PET/CT. They demonstrated that both PET-based texture analysis of dbPET and whole-body PET/CT had comparable classification power for the diagnosis of breast cancer (Satoh et al.). Fan et al. evaluated the value of texture analysis in the differential diagnosis of spinal metastases of ^{18}F-FDG PET/CT, and indicated that the combination of machine learning and texture parameters was more accurate than manual diagnosis. Zheng et al. applied a radiomics model in ^{18}F-FDG PET/CT to predict pathological mediastinal lymph node (pN) staging in patients with non-small cell lung cancer (NSCLC), and demonstrated an encouraging conclusion, suggesting that the pN staging and prediction have the potential to help with therapeutic planning. Apart from PET/CT, another nuclear medicine imaging method—single-photon emission computed tomography/computed tomography (SPECT/CT) also plays a role in the differentiation of benign and malignant tumors. Jin et al. investigated the feasibility of SPECT/CT images based on radiomics in differentiating bone metastases from benign bone lesions in patients with tumors. Both SPECT and SPECT/CT models showed better diagnostic accuracy than manual classification in the training and validation groups of patients diagnosed with vertebral bone metastases or benign bone lesions, indicating a new way in disease staging and treatment planning (Jin et al.).

In neurology, AI has also been applied to help for the distinguishment and mechanism research of neurodegenerative diseases. Xie et al. revealed a pattern of changes in β-amyloid (Aβ) deposition in cognitively normal healthy aging, which could be used to distinguish physiological changes from pathophysiological changes and help investigate the mechanism of Alzheimer's disease (AD). Zhou et al. presented a new deep learning model based on rate-distortion theory and an extreme learning machine model to distinguish AD, mild cognitive impairment (MCI), and normal controls (NC) in ^{18}F-AV45 PET/MR. This new deep learning model achieved higher accuracy, sensitivity, specificity, and area under curve (AUC) to separate AD, MCI, and NC groups than the previous models assessed (Zhou et al.).

The applications of AI are more than image processing, analysis, and interpretation, and also embrace searching for electronic health records, laboratory tests, and other information related to patients, which could assist physicians make optimal and personalized medical decisions for patients. With such abilities, AI provides more opportunities for assessing therapeutic responses and predicting survival rates. Tang et al. investigated the metabolic profiles of extratemporal in drug-resistant temporal lobe epilepsy (TLE) and efficiently predicted the surgery failure of TLE patients through PET, which could be used as predictive models for epilepsy surgery. Pinochet et al. evaluated the performance of a research prototype called PET Assisted Reporting System (PARS), which is based on a convolutional neural network, in clinical research. The PARS, determined total tumor metabolic volumes (TMTVs) on ^{18}F-FDG PET, was predictive of prognosis in patients with diffuse large B-cell lymphoma (DLBCL), but the evaluation efficacy was still needed to be enhanced and validated in miscellaneous cancers (Pinochet et al.). Yang et al. developed a radiomics score using the least absolute shrinkage and selection operator (LASSO) regression analysis in ^{18}F-FDG PET/CT-derived radiomic features, which proved to be a useful tool for predicting overall survival (OS) in adult hemophagocytic lymphohistiocytosis (HLH). Combining the radiomics score with the clinical parameters even performed better for predicting 6-month survival (Yang et al.). Apart from the above studies, AI can also assist to evaluate therapeutic responses and predict prognosis in other emerging treatments, including immunotherapy (8) and peptide radio receptor therapy (PRRT) (9).

Furthermore, with the powerful searching ability, AI can optimize the workflow and provide more detailed and organized information of patients, making it convenient for doctors to assess patient status more efficiently and precisely. Improtantly, AI plays a supportive role to relieve the physicians from labor-intensive but less cognitively demanding routine tasks, allowing them to focus on more mental work, such as patient care and image interpretation (1). Another step that is inseparable from AI in PET is imaging data processing, especially when it comes to standardizing imaging acquisition and reconstruction procedures. The repeatability of AI analysis in multicenter settings is essential to clinical translation. Zwanenburg performed a meta-analysis to evaluate the repeatability of PET imaging biomarkers (10). Based on the results, variations in the image acquisition, reconstruction, segmentation, and processing strongly affect the reliability of image biomarkers in models of different PET centers (10).

Although AI has tremendous potential in PET, it is critically important to be aware of its limitations. First, the reproducibility and reliability of AI algorithms are required. With the modeling becoming more and more complicated, the "black box" nature of AI makes it difficult to understand and explain the results of many AI models, especially in some DL models (11). The current trustworthy AI for health care prefers the explainability and stability of diverse and unknown data (12). Second, AI needs massive annotated data for learning and growing, which makes AI less reliable in small datasets. Currently, standard datasets, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) that is available to the public, are still relatively rare and in high demands. More open standard datasets will promote the development of AI. Last but not the least, there are still many ethical issues that need to be discussed, such as who is responsible for the legal and ethical issues if the AI diagnosis turns out to be wrong?

Despite the above limitations, it is obvious that AI plays a unique role in every step of PET, including patient information management, drug synthesis and administration,

image acquisition and processing, as well as report interpretation. AI can not only be used for helping make clinical decisions, but also as a research helper for discovering and investigating novel molecular biomarkers and their mechanisms in various diseases (13). With the powerful impact of AI in medical imaging, many physicians, especially radiologists, are concerned to be replaced by AI in the future. In fact, as Nensa et al. said, we should perceive the change as an opportunity rather than a threat (1). In the near future, or even at the moment, medical imaging physicians should not only simply focus on describing what images showing, but also need to pay more attention to all available information and data of patients, and conduct comprehensive analysis and interpretation to diseases diagnosis and therapeutic efficacy assessment or prediction, which greatly assist in providing more precise and personalized medical care for every individual patient.

AUTHOR CONTRIBUTIONS

HF wrote the manuscript. XL, KS, XW, and CZ edited the manuscript. All authors have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by the Opening Foundation of Hubei Key Laboratory of Molecular Imaging (2020fzyx003).

REFERENCES

1. Nensa F, Demircioglu A, Rischpler C. Artificial intelligence in nuclear medicine. *J Nucl Med.* (2019) 60:29S−37S. doi: 10.2967/jnumed.118.220590
2. Presotto L, Bettinardi V, De Bernardi E, Belli ML, Cattaneo GM, Broggi S, et al. PET textural features stability and pattern discrimination power for radiomics analysis: an "ad-hoc" phantoms study. *Phys Med.* (2018) 50:66–74. doi: 10.1016/j.ejmp.2018.05.024
3. Xu H, Lv W, Zhang H, Ma J, Zhao P, Lu L. Evaluation and optimization of radiomics features stability to respiratory motion in ^{18}F-FDG 3D PET imaging. *Med Phys.* (2021) 48:5165–78. doi: 10.1002/mp.15022
4. Seifert R, Weber M, Kocakavuk E, Rischpler C, Kersting D. Artificial intelligence and machine learning in nuclear medicine: future perspectives. *Semin Nucl Med.* (2021) 51:170–7. doi: 10.1053/j.semnuclmed.2020.08.003
5. Lee LIT, Kanthasamy S, Ayyalaraju RS, Ganatra R. The current state of artificial intelligence in medical imaging and nuclear medicine. *BJR Open.* (2019) 1:20190037. doi: 10.1259/bjro.20190037
6. Ninatti G, Kirienko M, Neri E, Sollini M, Chiti A. Imaging-based prediction of molecular therapy targets in NSCLC by radiogenomics and AI approaches: a systematic review. *Diagnostics.* (2020) 10:359. doi: 10.3390/diagnostics10060359
7. Lv Z, Fan J, Xu J, Wu F, Huang Q, Guo M, et al. Value of ^{18}F-FDG PET/CT for predicting EGFR mutations and positive ALK expression in patients with non-small cell lung cancer: a retrospective analysis of 849 Chinese patients. *Eur J Nucl Med Mol Imaging.* (2018) 45:735–750. doi: 10.1007/s00259-017-3885-z
8. Oriuchi N, Sugawara S, Shiga T. Positron emission tomography for response evaluation in microenvironment-targeted anti-cancer therapy. *Biomedicines.* (2020) 8:371. doi: 10.3390/biomedicines8090371
9. Roll W, Weckesser M, Seifert R, Bodei L, Rahbar K. Imaging and liquid biopsy in the prediction and evaluation of response to PRRT in neuroendocrine tumors: implications for patient management. *Eur J Nucl Med Mol Imaging.* (2021) 48:4016–27. doi: 10.1007/s00259-021-05359-3
10. Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur J Nucl Med Mol Imaging.* (2019) 46:2638–55. doi: 10.1007/s00259-019-04391-8
11. Rosenfeld A, Zemel R, Tsotsos JK. The Elephant in the Room. (2018). arXiv:1808.03305 [cs.CV].
12. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform.* (2021) 113:103655. doi: 10.1016/j.jbi.2020.103655
13. Mayerhoefer ME, Materka A, Langs G, Haggstrom I, Szczypinski P, Gibbs P, et al. Introduction to radiomics. *J Nucl Med.* (2020) 61:488–95. doi: 10.2967/jnumed.118.222893

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 14 December 2020
doi: 10.3389/fmed.2020.605002

FDG-PET Profiles of Extratemporal Metabolism as a Predictor of Surgical Failure in Temporal Lobe Epilepsy

Yongxiang Tang[1†], Guang Liao[1†], Jian Li[1], Tingting Long[1], Yulai Li[1], Li Feng[2], Dengming Chen[1], Beisha Tang[2,3] and Shuo Hu[1,3,4]*

[1] Department of Nuclear Medicine, Xiangya Hospital, Central South University, Changsha, China, [2] Department of Neurology, Xiangya Hospital, Central South University, Changsha, China, [3] National Clinical Research Center for Geriatric Diseases, Xiangya Hospital, Central South University, Changsha, China, [4] Key Laboratory of Biological Nanotechnology of National Health Commission, Xiangya Hospital, Central South University, Changsha, China

Objective: Metabolic abnormality in the extratemporal area on fluorine-18-fluorodeoxyglucose positron emission tomography (FDG-PET) is not an uncommon finding in drug-resistant temporal lobe epilepsy (TLE), however the correlation between extratemporal metabolic abnormalities and surgical long-term prognosis has not been fully elucidated. We aim to investigate FDG-PET extratemporal metabolic profiles predictive of failure in surgery for TLE patients.

Methods: Eighty-two patients with unilateral TLE (48 female, 34 male; 25.6 ± 10.6 years old; 37 left TLE, 45 right TLE) and 30 healthy age-matched controls were enrolled. Patients were classified either as experiencing seizure-recurrence (SZR, Engel class II through IV) or seizure-free (SZF, Engel class I) at least 1 year after surgery. Regional cerebral metabolism was evaluated by FDG-PET with statistical parametric mapping (SPM12). Abnormal metabolic profiles and patterns on FDG-PET in SZR group were evaluated and compared with those of healthy control and SZF subjects on SPM12. Volume and intensity as well as special brain areas of abnormal metabolism in temporal and extratemporal regions were quantified and visualized.

Results: With a median follow-up of 1.5 years, 60% of patients achieved Engel class I (SZF). SZR was associated with left TLE and widespread hypometabolism in FDG-PET visual assessment (*both* $p < 0.05$). All patients had hypometabolism in the ipsilateral temporal lobe but SZR was not correlated with volume or intensity of temporal hypometabolism (median, 1,456 vs. 1,040 mm^3; $p > 0.05$). SZR was correlated with extratemporal metabolic abnormalities that differed according to lateralization: in right TLE, SZR exhibited larger volume in extratemporal areas compared to SZF (median, 11,060 vs. 2,112 mm^3; $p < 0.05$). Surgical failure was characterized by Cingulum_Ant_R/L, Frontal_Inf_Orb_R abnormal metabolism in extratemporal regions. In left TLE, SZR presented a larger involvement of extratemporal areas similar to right TLE but with no significant (median, 5,873 vs. 3,464 mm^3; $p > 0.05$), Cingulum_Ant_ R/L, Parietal_Inf_L, Postcentral_L, and Precuneus_R involved metabolic abnormalities were correlated with SZR.

OPEN ACCESS

Edited by:
Kuangyu Shi,
University of Bern, Switzerland

Reviewed by:
Salvatore Annunziata,
Catholic University of the Sacred Heart, Italy
Zhen Cheng,
Stanford University, United States

****Correspondence:***
Shuo Hu
hushuo2018@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: *11 September 2020*
Accepted: *23 November 2020*
Published: *14 December 2020*

Citation:
Tang Y, Liao G, Li J, Long T, Li Y, Feng L, Chen D, Tang B and Hu S (2020) FDG-PET Profiles of Extratemporal Metabolism as a Predictor of Surgical Failure in Temporal Lobe Epilepsy. Front. Med. 7:605002. doi: 10.3389/fmed.2020.605002

Conclusions: Extratemporal metabolic profiles detected by FDG-PET may indicate a prominent cause of TLE surgery failure and should be considered in predictive models for epilepsy surgery. Seizure control after surgery might be improved by investigating extratemporal areas as candidates for resection or neuromodulation.

Keywords: PET in epilepsy, prognosis, epilepsy, image processing, PET

INTRODUCTION

The goal of epilepsy surgery is to render the patient seizure-free. Surgical treatment of drug-resistant temporal lobe epilepsy (TLE) has proven superior to medical management of the disease (1). However, the proportion of seizure-free following TLE surgery remains suboptimal (2, 3). In ∼50% of cases in which surgery fails to achieve seizure freedom, patients who continue to experience seizures after surgery are directly associated with an even lower quality of life (4). It has proven difficult to identify from existing literature independent biomarker highly predictive of TLE surgery failure. Therefore, understanding the prominent reasons for surgical failure and identifying effective indicators to facilitate early evaluation remain of paramount importance in the context of epilepsy care (3, 5–7).

Evidence from neuroimaging and electrophysiology studies has consistently shown that epilepsy is a disease affecting neural networks, with abnormalities occurring well-beyond the locus of ictogenesis (8). While recurrent seizures following surgery suggested that some epileptogenic tissue distinct from the primary temporal lobe epileptogenic zone has not been resected, and those areas might reflect dual pathology beyond seizure onset zone, but usually negative in magnetic resonance imaging (MRI) or impossibility of whole brain electrocorticogram (ECoG) coverage (5, 7), thus posing a difficult diagnostic challenge. Emerging evidence suggests that temporal lobe hypometabolism on fluorine-18-fluorodeoxyglucose positron emission tomography (FDG-PET) can provide relevant information on the epileptogenic zone extent and surgical outcome (9–11). The highest clinical benefit of FDG-PET can be achieved in patients with MRI-negative TLE (12). However, metabolic abnormality in the extratemporal cortex on FDG-PET is not an uncommon finding and seems to be associated with surgical outcome in TLE (13, 14), but the correlation between extratemporal metabolic abnormalities and surgical long-term prognosis has not been fully elucidated. Moreover, it is not clear which profiles of temporal or extratemporal metabolic abnormalities are more important for TLE surgical failure.

In this study, we analyzed patients with unilateral TLE who had undergone identical surgical resections. Long-term seizure outcomes were analyzed to determine the potential usefulness of FDG-PET temporal and extratemporal metabolic intensity, volume and/or specific brain areas for predicting postoperative seizure recurrence in TLE patients. Imaging processing was performed with statistical parametric mapping (SPM 12).

Abbreviations: AAL, automated anatomical labeling; AMTR, anteromedial temporal resection; EEG, electroencephalogram; FDG-PET, fluorine-18 fluorodeoxyglucose positron emission tomography; HS, hippocampal sclerosis; ILAE, International League Against Epilepsy; IQR, interquartile range; LTLE, left temporal lobe epilepsy; MNI, Montreal Neurological Institute; MRI, magnetic resonance imaging; ROC, receiver operating characteristic; RTLE, right temporal lobe epilepsy; SD, standard deviation; SPECT, single photon emission computed tomography; SPM 12, statistical parametric mapping; SZF, seizure free; SZR, seizure recurrence; TLE, temporal lobe epilepsy; TPE, temporal plus epilepsy.

MATERIALS AND METHODS

Patients and Healthy Controls

We retrospectively reviewed 82 drug-resistant unilateral TLE patients (45 right TLE [RTLE], 37 left TLE [LTLE]) who had received preoperative FDG-PET between April 2014 and April 2018. Diagnosis of drug-resistant unilateral TLE was based upon comprehensive clinical assessment and criteria of the International League Against Epilepsy (ILAE) (15). Each patient was surgically treated by identical anteromedial temporal resection (AMTR) without extratemporal resections as described by Spencer et al. (16). Pathology was assessed from postoperative pathology reports. The determination of postsurgical outcome was based on in-person interviews and patient assessment during clinic follow-up. Patients without 1-year follow-up were excluded from analysis.

Thirty healthy age-matched volunteers were recruited as normal controls. None had a history of head injury or a major neurological, physical, or psychiatric disorder, including drug and alcohol misuse.

Ethical Approval and Patient Consent

Study protocols was approved by the Ethical Commission of Medical Research Involving Human Subjects at Region of Xiangya Hospital, Central South University, China [IBR{C}NO. (201412455)]. Written informed consent was provided by all participants (patients and controls) in accordance with the Helsinki Declaration.

Clinical Data

In addition to undergoing FDG-PET, patients completed a pre-surgical assessment consisting of a detailed clinical history and examination, video electroencephalogram (EEG) monitoring, brain MRI, neuropsychiatric testing and intracranial EEG monitoring when indicated. FDG-PET classification was defined as previously described (17) and classified into two subtypes according to the visual assessment: focal or widespread hypometabolism.

Outcome assessments were performed 3 and 12 months after surgery and at yearly intervals thereafter. All patients were interviewed in detail for seizure recurrence, if any, and date of recurrence. Surgical outcomes were classified based on the Engel

Surgical Outcome scale (18, 19) as either seizure-free (SZF; Engel class I) or seizure-recurrence (SZR; Engel class II through IV).

FDG-PET Image Acquisition and Processing

FDG-PET was acquired using a Discovery Elite PET/CT scanner (GE Healthcare) prior to surgical resection. FDG was injected intravenously at a mean dose of 148 MBq. Images were acquired in three dimensions over a 60 min time period, following scanning protocol described by Tang et al. (11). Image processing was performed using the SPM12 (Wellcome Department of Cognitive Neurology, London, UK). Individual FDG-PET image volumes were spatially normalized into standard stereotactic Montreal Neurological Institute (MNI) space with voxel sizes of 2×2×2. An 8-mm full-width-half-maximum Gaussian kernel was used to improve between-participant spatial alignment and smooth data for statistical analysis. The right and left hemispheres were analyzed separately to detect lateralization effects on surgical outcomes (20). Image intensity between participants was normalized to prevent interparticipant variability in cerebral tracer uptake from masking regional changes. Increased or decreased metabolism was considered statistically significant when uncorrected p = 0.001 with cluster level above 20 contiguous voxels. After data preprocessing using SPM significant clusters were visualized, reported and anatomically labeled using the xjView (http://www.alivelearn.net/xjview), REST and BrainNet Viewer Toolkit (21, 22). Data include metabolic profile information about the clusters, including number of voxels (or volumes), anatomical term of Automated Anatomical Labeling (AAL) areas and peak intensity of each cluster.

Statistical Analysis

All data were analyzed using SPSS software (IBM SPSS Statistics, Version 18.0). Numerical data are presented as mean ± *SD* or Median (IQR). Student *t*-tests or Mann-Witney tests and Pearson χ^2 test were used for continuous and categorical variables in between-group comparisons, as appropriate. For FDG-PET image SPM analysis, the general linear model was used to carry out the appropriate voxel-by-voxel univariate statistical tests. Individual TLE SPM analysis was firstly performed using 30 healthy controls. The volume of metabolism changing in temporal, extratemporal areas and whole brain for each patient was then calculated as well as peak intensity for the ipsilateral temporal lobe. Abnormal metabolic volume in RTLE and LTLE was compared between two outcome groups using the Mann-Whitney test. Then, individual SZR SPM analysis was performed using SZF group, increased and decreased metabolism of extratemporal brain area were visualized and frequency was calculated. We then compared baseline glucose uptake values of each outcome group (SZR and SZF separately) and healthy controls in RTLE and LTLE using an analysis of covariance (ANCOVA) with group as the between-subject factor and age and sex as confounding covariates (23, 24). A two-sample *t*-test was used to compared the different groups. Other comparisons were performed between two outcome groups in both sides of TLE. Statistical significance was defined as $p < 0.05$.

RESULTS

Clinical Data

Eighty-two refractory TLE patients (48 female, 59%) met the inclusion criteria of isolated AMTR with 1 or more years of follow-up. Median follow-up time was 1.5 years (IQR 1.2–2.3) with a maximum follow-up time of 5 years. Patient clinical characteristics are shown in **Table 1**. Briefly, 49 of 82 patients (60%) obtained an Engle class I outcome. SZF was more frequently obtained in RTLE (73%) than in LTLE (43%) and the difference was significant ($p = 0.006$). The main clinical variable differing significantly between SZF and SZR outcome groups were seen for focal or widespread hypometabolism on FDG-PET in RTLE. SZF group patients exhibited a narrower range of hypometabolism in FDG-PET analysis by visual assessment in RTLE (85.2% at focal hypometabolism vs. 55.6% widespread hypometabolism, $p = 0.028$). Regardless of whether there was hippocampal sclerosis(HS) in the postoperative histopathology or MRI, there was no difference in their surgical outcomes (*both* $p > 0.05$).

Volume of Metabolic Abnormalities in Temporal and Extratemporal Areas

All patients exhibited hypometabolism in the temporal cortex ipsilateral to the epileptogenic region (**Figure 1**). Hypometabolism predominated in the temporal lobe in the SZF cohort, with some patients also exhibiting minimal hypometabolism or hypermetabolism outside the temporal lobe. By contrast, larger differences were observed in extratemporal areas in SZR cohort patients. Compared to healthy controls, volume of metabolic abnormalities measured in temporal areas, extratemporal areas and whole brain have differently correlated with surgical outcome, and relative to lateralization of TLE. In RTLE, relative to the SZF cohort, SZR cohort showed larger volume metabolic abnormalities in extratemporal areas (median, 11,060 vs. 2,112 mm^3; $p = 0.02$). The metabolic abnormality volumes for temporal lobe and whole brain were larger in SZR (median, 3,356 and 14,252 mm^3) than in SZF cases (median, 936 and 4,200 mm^3) but the differences were not significant (*both* $p > 0.05$). In LTLE, the same trend was seen as in RTLE patients but no significant volume differences were found in temporal lobe, extratemporal areas and whole brain between SZR and SZF abnormal cerebral metabolism (temporal lobe: 1,056 vs. 1,040 mm^3; extratemporal areas: 5,872 vs. 2,560 mm^3; whole brain: 5,872 vs. 3,464 mm^3; *all* $p > 0.05$). These results were in part consistent with focal or widespread hypometabolism classification by visual assessment on FDG-PET. No significant difference in volume of ipsilateral temporal hypometabolism was observed in both sides TLE (Mann-Witney test, $p > 0.05$) (**Table 1**). Those suggest that hypometabolism volume in TLE foci had no effect on surgical outcome, while the volume of extratemporal metabolic abnormalities could affect surgical outcome especially in RTLE. Larger volume of extratemporal metabolic change correlated with worse surgical prognosis. The critical volume of metabolic change outside RTLE foci (calculated from ROC curve) was 12,580 mm^3.

TABLE 1 | Patient clinical characteristics and surgical outcomes.

	LTLE patients ($n = 37$)			RTLE patients ($n = 45$)			All patients ($n = 82$)		
	SZF	**SZR**	***p***	**SZF**	**SZR**	***P***	**SZF**	**SZR**	***p***
Gender (male/female), *n*	10/6	11/10	NS	20/13	7/5	NS	30/19	18/15	NS
Age (mean ± *SD*), y	24.8 ± 8.2	27.0 ± 12.1	NS	25.8 ± 11.2	23.8 ± 9.7	NS	25.5 ± 10.3	25.8 ± 11.3	NS
Age at onset (mean ± *SD*), y	10.7 ± 7.7	11.7 ± 8.3	NS	14.9 ± 11.1	9.7 ± 6.9	NS	13.5 ± 10.2	11.0 ± 7.8	NS
Duration of epilepsy (mean ± *SD*), y	14.4 ± 8.9	15.2 ± 8.6	NS	11.0 ± 7.7	14.2 ± 7.6	NS	12.1 ± 8.2	14.8 ± 8.1	NS
Surgicalside (L/R)	16 (43%)	21 (57%)	-	33 (73%)	12 (27%)	-	16/33(60%)	2112 (40%)	**$p < 0.05$**
Histopathology, *n* (%)			NS			NS			NS
HS	10 (27)	12 (32)		21 (47)	5 (11)		31 (38)	17 (21)	
Non-HS	6 (16)	9 (25)		12 (27)	7 (15)		18 (21)	16 (20)	
Handedness (L/R)	1/15	0/21	NS	1/32	0/12	NS	2/47	0/33	NS
History, *n* (%)			NS			NS			NS
Febrile seizures.	4 (11)	2 (5)		4 (9)	2 (5)		8 (10)	4 (5)	
Brain injury	1 (3)	4 (11)		1 (2)	0 (0)		2 (2)	4 (5)	
Without	10 (27)	15 (40)		24 (53)	9 (20)		34 (41)	24 (29)	
Encephalitis	1 (3)	0 (0)		4 (9)	1 (2)		5 (7)	1 (1)	
Psychiatric complication (with/without)	1/15	1/20	NS	0/33	1/11	NS	1/48	2/31	NS
Arua (with/without)	7/9	12/9	NS	18/15	6/6	NS	25/24	18/15	NS
Family history of epilepsy (with/without)	0/16	0/21	NS	0/33	1/11	NS	0/49	1/32	NS
Visual evaluation of PET (local/wide)	7/9	8/13	NS	23/10	4/8	**0.028**	30/19	12/21	**0.027**
Result of MRI (positive/negative)	13/3	14/7	NS	28/5	10/2	NS	41/8	24/9	NS
Volume of metabolic change									
Temporal areas, median (IQR)	1,040 (62–4,428)	1,056 (0–7,328)	NS	936 (52–5,076)	3,356 (64–11,000)	NS	1,040 (52–4,896)	1,456 (0–9580)	NS
Extratemporal areas, median (IQR)	3,464 (1,048–11,502)	5,872 (672–30,140)	NS	2,112 (176–36,284)	11,060 (930–22,290)	**$p < 0.05$**	2,112 (176–6,708)	8,424 (732–23,884)	**$p < 0.05$**

p-value is derived from the univariable association analyses between each of the clinicopathologic variables and surgical outcome.
HS, hippocampal sclerosis; LTLE, left temporal lobe epilepsy; MRI, magnetic resonance imaging; NS, not significant; PET, positron emission tomography; RTLE, right temporal lobe epilepsy; SD, standard deviation; SZF, seizure-free (Engel class I); SZR, seizure recurrence (Engel class II through IV).
Bold values indicates significant difference ($p < 0.05$).

Intensity of Metabolic Abnormalities in Temporal and Extratemporal Areas

Compared to healthy controls, hypometabolism was predominant in the ipsilateral temporal lobe (mesial, lateral and polemesial, lateral and pole). Peak intensity of hypometabolism in RTLE was −7.61 for SZR, −7.14 for SZF. Values of LTLE were −6.78 for SZR and −7.98 for SZF. When comparing SZR to SZF groups, differences in temporal lobe area were insufficient for detection on SPM images in either RTLE or LTLE even if the threshold was adjusted to $p = 0.005$ (**Figure 2**).

In addition to hypometabolism in temporal lobe foci, SZR cohort might show multiple hypo- or hyper-metabolism areas outside the temporal lobe. Compared to healthy controls, LTLE patients exhibited the most severe abnormalities in Parietal_Inf_L, Postcentral_L (hypometabolism, peak intensity −5.16, −4.58) and Cingulum_Ant_R, Precuneus_R (hypermetabolism, peak intensity 5.74, 4.62) in the SZR cohort. In RTLE patients, the greatest discrepancy was found in the Frontal_Inf_Orb_R (hypometabolism, peak intensity −3.67) and Cingulum_Ant_R/L (hypermetabolism, peak intensity 4.88, 4.17) in the SZR cohort (**Figure 1**).

TLE patients had hypometabolism in the temporal cortex ipsilateral to the epileptogenic region, but surgical outcomes were not associated with volume and intensity of ipsilateral temporal hypometabolism. SZR correlated with extratemporal metabolic profile that differed according to TLE lateralization.

Special Extratemporal Brain Areas of Metabolic Abnormalities

Comparing each SZR patient to the SZF cohort revealed the frequency of involvement of each extratemporal region of all SZR patients. In both LTLE and RTLE, the frequencies of bilateral frontal lobe and left parietal lobe involvement figured in the top three areas. Other areas included cingulate gyrus, bilateral occipital lobe, contralateral temporal lobe, insula, and caudate nucleus (**Figure 3**). However, comparing glucose uptake values between the two surgical outcome cohorts, we found that the peak intensity of metabolic changes in SZR patients differed with respect to comparisons with healthy controls. SZR vs. SZF differences were not strong enough to be visualized on SPM images at the fixed threshold. Lowering the threshold to $p = 0.005$ revealed

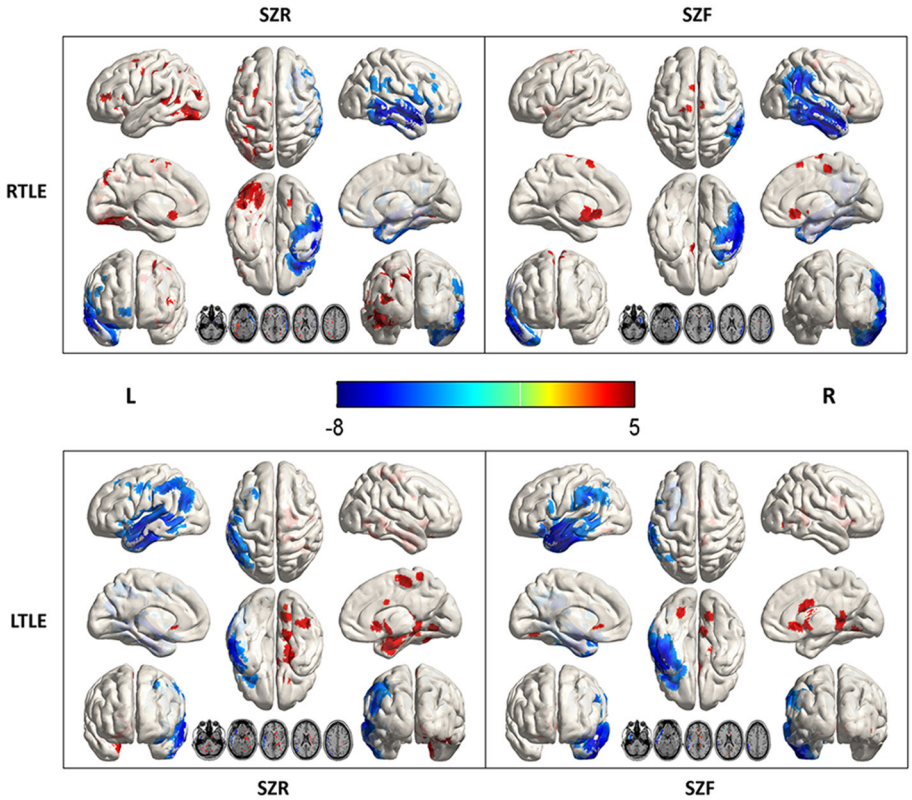


FIGURE 1 | Metabolic data from 45 right (R) and 37 left (L) temporal lobe epilepsy (TLE) according to surgical outcome. Comparison with 30 healthy controls (SPM12, $p < 0.001$). In RTLE, hypometabolism (blue) was predominant in the temporal cortex ipsilateral to the epileptogenic region, with mild extratemporal involvement in the seizure-free group (SZF, Engel class I). In contrast, a larger volume involvement of nearby brain areas of foci and other perisylvian regions, as well as the Frontal_Inf_Orb_R and Cingulum_Ant_R/L, were found in the seizure recurrence group (SZR, Engel class II through IV). In LTLE, the difference between SZR and SZF outcome groups was much less marked. Parietal_Inf_L, Postcentral_L (hypometabolism, blue), and Cingulum_Ant_R, Precuneus_R (hypermetabolism, red) associated with an unfavorable outcome.

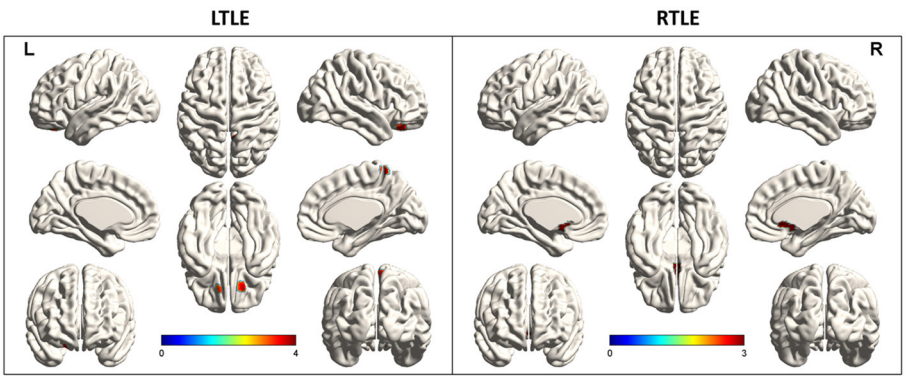


FIGURE 2 | Metabolic differences according to outcome in right (R) and left (L) temporal lobe epilepsy (TLE) ($p < 0.005$). Areas of significant hypermetabolism in seizure recurrence (SZR) compared to seizure-free (SZF) differed between RTLE and LTLE. Differences in temporal lobe and extratemporal region were not strong enough to be visualized on SPM images in both RTLE and LTLE. Differences in temporal lobe area were insufficient for detection on SPM images in either RTLE or LTLE even when the threshold was lowered to $p = 0.005$. Hypermetabolism (red) in Precuneus_R, Frontal_Sup_Orb_R/L of LTLE and Cingulum_Ant_L/R of RTLE was evident when the threshold was set to $p = 0.005$.

LTLE hypermetabolism in Precuneus_R, Frontal_Sup_Orb_R/L, and RTLE hypermetabolism in Cingulum_Ant_L/R, but the difference in temporal lobe area was insufficient for detection (**Figure 2**).

DISCUSSION

Here we describe a large group of patients with a homogeneous clinicopathologic syndrome undergoing standard AMTR for TLE at a single center with a long postoperative follow-up. Metabolic abnormalities both in temporal and extratemporal areas were visualized and quantified using FDG-PET. We showed that the volume and intensity of metabolic changes in extratemporal areas and special brain areas assessed by FDG-PET correlated with a substantial proportion of surgical failures. We went further, to explore surgical outcomes in TLE with specific combinations of clinical characteristics. Hemispheric asymmetry was found to correlate with surgical results. Histopathology and MRI assessment of HS, as previously described (14, 25), may have little to do with the prognosis of surgery.

At the epilepsy center where this study was undertaken, an MRI normal hippocampus is an indication for FDG-PET evaluation even when other non-invasive evaluations lateralize to one temporal lobe foci. In this study, for some cases in which ictal onset localized to temporal lobe and an AMTR was performed, the neuronal loss in non-HS was milder than typically observed for mesial temporal sclerosis, but it does not indicate a higher rate of post-operation seizure control (26), on the contrary, sometimes negative MRI could prevent epilepsy surgeons from discovering temporal lobe epileptic foci. In TLE that is resistant to surgery, our glucose metabolism study has suggested an epileptogenic involvement of cortical areas outside temporal structures, both in cases with HS and non-HS. Accordingly, we showed that the volume and intensity of metabolic changes in temporal lobe foci had no effect on surgical outcomes. Previously, TLE patients with greater maximal temporal asymmetries were found to be less likely to achieve seizure-free status (27), but this analysis did not consider effects of epileptic foci and other brain regions on the contralateral temporal lobe, resulting in a reduction of the asymmetry index. Altogether, our results and other multivariate studies suggest that ipsilateral temporal hypometabolism may not be a prognostic factor, but rather a diagnostic indicator (13, 28).

The reasons for surgery failure are complex and variable, but emerging evidence points to failure to resect epileptogenic areas either within or outside the operated temporal lobe (7, 29–31). This situation can be defined as a dual pathology combining an extratemporal epileptogenic lesion and temporal epileptogenic zone (32, 33). Our results may be better explained by attributing epileptogenic potential to sites of metabolic abnormalities in extratemporal areas that were not included in resection. Several studies have reported extratemporal involvement associated with poor postoperative outcome, including nearby structures outside the standard resective margins or distant neocortical areas (13, 14, 25, 34). We found worse postoperative outcomes, especially among RTLE patients, correlated with larger volume or visual assessment range of metabolic changes in extratemporal areas detected by FDG-PET imaging.

Surgical failures in TLE have pointed to a subset of patients with primary seizure onset in the temporal lobe plus an "epileptogenic zone" that extends to nearby structures outside the standard resective margins, termed temporal-plus epilepsy (TPE) (5, 7, 34). Generally, extra-temporal targets are selected on the basis of alternative hypotheses formulated regarding the location(s) and extent of the epileptogenic zone(s) (35). The most frequently investigated brain regions are the temporo-parieto-occipital junction, fronto-basal and orbito-frontal cortex, suprasylvian operculum, and insula. Similarities in seizure recurrence rates as well as visual assessment of FDG-PET images between the extensive hypometabolism described in this study and TPE subgroups (7) suggest the same factors behind failed RTLE surgery. Except for nearby brain areas of foci, however, almost all other cortical areas could be targeted (34). In addition, for LTLE we found that while FDG-PET visual assessment and volume of metabolic change in extratemporal areas did not predict surgical outcomes, special cortical metabolic change in extratemporal areas could be predictive. Outcome predictions for LTLE patients were more challenging than for RTLE patients, related to ranges of predictive factors and this may account for the hemispheric asymmetry in metabolic patterns with respect to outcomes.

Such an explanation for surgical failure implies that a distributed epileptogenic network rather than a single epileptogenic focus may underlie surgically refractory epilepsy. Identification of epileptogenic foci and influences on networked extra-focal areas responsible for continued or recurrent postoperative seizures remain at the forefront of research to improve outcomes (36, 37). In addition to larger volume metabolic abnormalities in extratemporal areas linked to unfavorable outcomes, we describe special patterns associated with TLE surgical failure ranging from the frontal, parietal to occipital lobe and contralateral networks; some of which are consistent with electroclinical patterns corresponding to anterior and posterior spread (25). The present study's data support potential network theory of epileptogenicity in which nodes outside the seizure onset zone are implicated in seizure generation (37, 38). In TLE, the profile of metabolic abnormalities in extratemporal areas including volume, intensity and involvement of special brain areas assessed by FDG-PET, may indicate that these extratemporal limbic nodes are critical for epileptogenic network onset (39). Going forward, these should be sampled when an intracranial electroencephalography study of presumed temporal onset is performed. A shift toward treating the seizure network rather than a seizure-onset focus could alter the surgical and neuromodulatory management of focal epilepsy and guide electrode placement.

To address this issue therapeutically, extending resection to include nearby structures is one possible approach, when the metabolic pattern includes dispensable cortical areas such as the non-dominant lateral temporal cortex. When extended resection is not possible (e.g., dominant temporal lobe, frontal neocortex, or cerebral cortex motor area), treatment might involve placement of responsive neurostimulation electrodes at

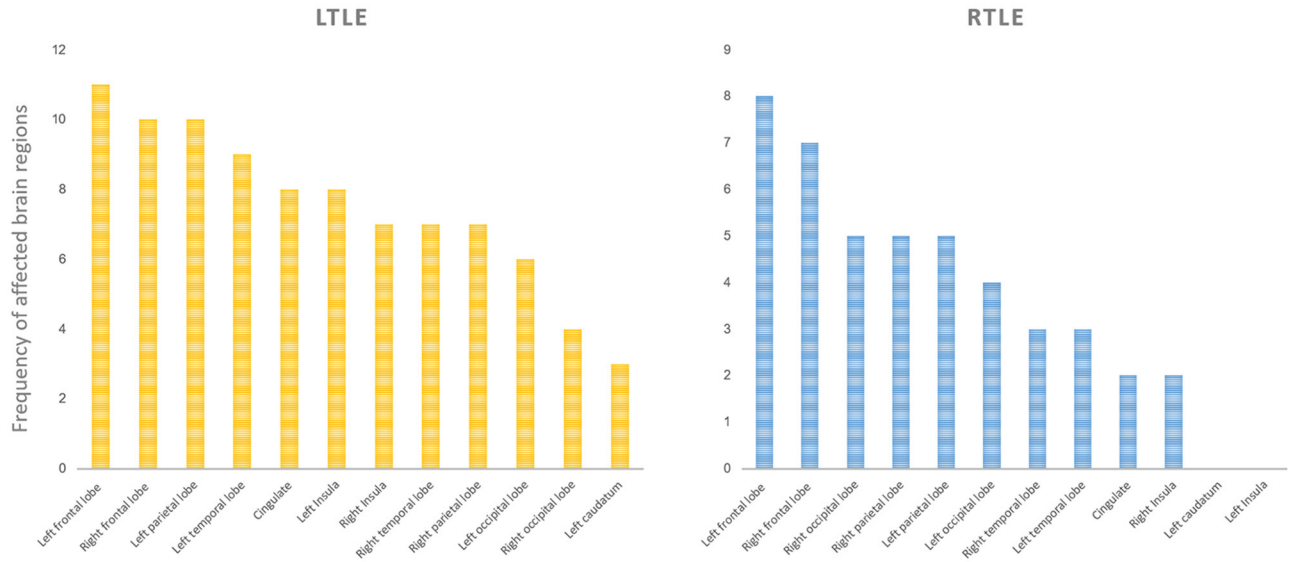


FIGURE 3 | Involvement frequency of extratemporal region in all seizure recurrence (SZR) patients compared to the seizure-free (SZF) group. We compared the frequency of involvement of each extratemporal region of all patients in the SZR group. In both right (R) and left (L) temporal lobe epilepsy (TLE), the bilateral frontal lobe and left parietal lobe involvement frequencies were in the top three. Other areas included cingulate gyrus, bilateral occipital lobe, contralateral temporal lobe, insula, and caudate nucleus.

these extratemporal limbic nodes. Responsive neurostimulation electrodes would allow long-term recordings to validate nodes that are responsible for recurrent seizures as well as determine whether these sites are responsive to neuromodulation (40). When extratemporal areas such as the parietal motor area or occipital lobe (RTLE: Cingulum_Ant_R/L, Frontal_Inf_Orb_R; LTLE: Cingulum_Ant_L/R, Parietal_Inf_L, Postcentral_L, and Precuneus_R) are identified as metabolic change network nodes, it is essential to carefully consider possible adverse outcomes of resection, ablation or neurostimulation. While AMTR may likely fail for such patients, the effectiveness of possible alternative approaches is not yet clear. Advanced functional neuroimaging techniques have identified different phenotypes within the TLE spectrum, and specific metabolic patterns have poor surgical outcomes (41). Both extended resection and responsive neurostimulation treatment approaches will require FDG-PET neuroimaging for treatment planning and clear definition of projected network nodes.

This study had several limitations. First, this was a retrospective analysis and all patients underwent FDG-PET scan, leading to a low proportion of class I Engel outcomes (60%) especially among LTLE patients (43%). As patients without FDG-PET imaging were not included, which might have led to some selection and ascertainment biases. The decision to undergo surgery for epilepsy is complex, and involves consideration of the patient's baseline disease burden and overall clinical picture. It is not solely based on the probability of achieving freedom from seizures. Our findings are not meant to replace clinical judgment, but rather to assist decision making by providing an objective estimate of one key decision-driving factor-postoperative seizure-recurrence. Finally, preoperative video-EEG data, structural MRI neuroimaging and results of more sophisticated diagnostic tests such as ictal SPECT and invasive electroencephalogram were not analyzed in study patients who failed TLE surgery.

In conclusion, Extratemporal metabolic profiles detected by FDG-PET, in particular volume and intensity and affected special extratemporal brain areas may be associated with unfavorable postoperative seizure outcome in TLE and should therefore be considered in predictive models for epilepsy surgery. For TLE confirmed to have specific extratemporal metabolic abnormalities on FDG-PET, AMTR appears very unlikely to control seizures and should not be advised.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The dataset used and analyzed for the current study is available from the corresponding author on reasonable request. Requests to access these datasets should be directed to Shuo Hu, hushuo2018@163.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Commission of Medical Research Involving Human Subjects at Region of Xiangya Hospital, Central South University, China. Written informed consent to participate in

this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

YT and GL designed the method, acquisition of data, and prepared the manuscript. SH, LF, and BT designed the method, aided in data analysis, revised, and approved the manuscript. JL, TL, YL, and DC aided in data acquisition and interpretation. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by grant no. 81801740 from National Natural Science Foundation of China, grant no. 2020JJ5922 from Natural Science Foundation of Hunan Province, and grant no. kq1901125 from Science and Technology Project of Changsha.

ACKNOWLEDGMENTS

The authors extend their deepest appreciation to the patient volunteers and their families. We thank the staff at the XiangYa Hospital neurosurgery for their expert assistance in this work, SPM12 (SPM12, Wellcome Department of Cognitive Neurology, London, UK), xjView, REST, and BrainNet Viewer MATLAB toolboxes for freely providing software analysis tools.

REFERENCES

1. Tellez-Zenteno JF, Dhar R, Wiebe S. Long-term seizure outcomes following epilepsy surgery: a systematic review and meta-analysis. *Brain.* (2005) 128:1188–98. doi: 10.1093/brain/awh449
2. Dwivedi R, Ramanujam B, Chandra PS, Sapra S, Gulati S, Kalaivani M, et al. Surgery for drug-resistant epilepsy in children. *N Engl J Med.* (2017) 377:1639–47. doi: 10.1056/NEJMoa1615335
3. West S, Nevitt SJ, Cotton J, Gandhi S, Weston J, Sudan A, et al. Surgery for epilepsy. *Cochrane Database Syst Rev.* (2019) 6:Cd010541. doi: 10.1002/14651858.CD010541.pub3
4. Sajobi TT, Jette N, Fiest KM, Patten SB, Engbers JD, Lowerison MW, et al. Correlates of disability related to seizures in persons with epilepsy. *Epilepsia.* (2015) 56:1463–9. doi: 10.1111/epi.13102
5. Harroud A, Bouthillier A, Weil AG, Nguyen DK. Temporal lobe epilepsy surgery failures: a review. *Epilepsy Res Treat.* (2012) 2012:201651. doi: 10.1155/2012/201651
6. Jehi L, Yardi R, Chagin K, Tassi L, Russo GL, Worrell G, et al. Development and validation of nomograms to provide individualised predictions of seizure outcomes after epilepsy surgery: a retrospective analysis. *Lancet Neurol.* (2015) 14:283–90. doi: 10.1016/S1474-4422(14)70325-4
7. Barba C, Rheims S, Minotti L, Guenot M, Hoffmann D, Chabardes S, et al. Temporal plus epilepsy is a major determinant of temporal lobe surgery failures. *Brain.* (2016) 139:444–51. doi: 10.1093/brain/awv372
8. Gleichgerrcht E, Kocher M, Bonilha L. Connectomics and graph theory analyses: novel insights into network abnormalities in epilepsy. *Epilepsia.* (2015) 56:1660–8. doi: 10.1111/epi.13133
9. Muhlhofer W, Tan YL, Mueller SG, Knowlton R. MRI-negative temporal lobe epilepsy-what do we know? *Epilepsia.* (2017) 58:727–42. doi: 10.1111/epi.13699
10. Sidhu MK, Duncan JS, Sander JW. Neuroimaging in epilepsy. *Curr Opin Neurol.* (2018) 31:371–8. doi: 10.1097/WCO.0000000000000568
11. Tang Y, Liow JS, Zhang Z, Li J, Long T, Li Y, et al. The evaluation of dynamic FDG-PET for detecting epileptic foci and analyzing reduced glucose phosphorylation in refractory epilepsy. *Front Neurosci.* (2018) 12:993. doi: 10.3389/fnins.2018.00993
12. Beaurain M, Salabert AS, Ribeiro MJ, Arlicot N, Damier P, Le Jeune F, et al. Innovative molecular imaging for clinical research, therapeutic stratification, and nosography in neuroscience. *Front Med.* (2019) 6:268. doi: 10.3389/fmed.2019.00268
13. Choi JY, Kim SJ, Hong SB, Seo DW, Hong SC, Kim BT, et al. Extratemporal hypometabolism on FDG PET in temporal lobe epilepsy as a predictor of seizure outcome after temporal lobectomy. *Eur J Nucl Med Mol Imaging.* (2003) 30:581–7. doi: 10.1007/s00259-002-1079-8
14. Chassoux F, Artiges E, Semah F, Laurent A, Landre E, Turak B, et al. (18)F-FDG-PET patterns of surgical success and failure in mesial temporal lobe epilepsy. *Neurology.* (2017) 88:1045–53. doi: 10.1212/WNL.0000000000003714
15. Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, van Emde Boas W, et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. *Epilepsia.* (2010) 51:676–85. doi: 10.1111/j.1528-1167.2010.02522.x
16. Spencer DD, Spencer SS, Mattson RH, Williamson PD, Novelly RA. Access to the posterior medial temporal lobe structures in the surgical treatment of temporal lobe epilepsy. *Neurosurgery.* (1984) 15:667–71. doi: 10.1227/00006123-198411000-00005
17. Malmgren K, Thom M. Hippocampal sclerosis–origins and imaging. *Epilepsia.* (2012) 53(Suppl.4):19–33. doi: 10.1111/j.1528-1167.2012.03610.x
18. Engel JJr, Wiebe S, French J, Sperling M, Williamson P, Spencer D, et al. Practice parameter: temporal lobe and localized neocortical resections for epilepsy. *Epilepsia.* (2003) 44:741–51. doi: 10.1046/j.1528-1157.2003.48202.x
19. Engel JJr, Wiebe S, French J, Sperling M, Williamson P, Spencer D, et al. Practice parameter: temporal lobe and localized neocortical resections for epilepsy: report of the Quality Standards Subcommittee of the American Academy of Neurology in association with the American Epilepsy Society and the American Association of Neurological Surgeons. *Neurology.* (2003) 60:538–47. doi: 10.1212/01.WNL.0000055086.35806.2D
20. Chen D, Jiang J, Lu J, Wu P, Zhang H, Zuo C, et al. Brain network and abnormal hemispheric asymmetry analyses to explore the marginal differences in glucose metabolic distributions among Alzheimer's disease, Parkinson's disease dementia, and lewy body dementia. *Front Neurol.* (2019) 10:369. doi: 10.3389/fneur.2019.00369
21. Song XW, Dong ZY, Long XY, Li SF, Zuo XN, Zhu CZ, et al. REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS ONE.* (2011) 6:e25031. doi: 10.1371/journal.pone.0025031
22. Xia M, Wang J, He Y. BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS ONE.* (2013) 8:e68910. doi: 10.1371/journal.pone.0068910
23. Zhang H, Wu P, Ziegler SI, Guan Y, Wang Y, Ge J, et al. Data-driven identification of intensity normalization region based on longitudinal coherency of (18)F-FDG metabolism in the healthy brain. *Neuroimage.* (2017) 146:589–99. doi: 10.1016/j.neuroimage.2016.09.031
24. Jiang J, Sun Y, Zhou H, Li S, Huang Z, Wu P, et al. Study of the influence of age in (18)F-FDG PET images using a data-driven approach and its evaluation in Alzheimer's disease. *Contrast Media Mol Imaging.* (2018) 2018:3786083. doi: 10.1155/2018/3786083
25. Chassoux F, Artiges E, Semah F, Desarnaud S, Laurent A, Landre E, et al. Determinants of brain metabolism changes in mesial temporal lobe epilepsy. *Epilepsia.* (2016) 57:907–19. doi: 10.1111/epi.13377
26. Spencer DD, Spencer SS. Hippocampal resections and the use of human tissue in defining temporal lobe epilepsy syndromes. *Hippocampus.* (1994) 4:243–9. doi: 10.1002/hipo.450040303

27. Lin TW, de Aburto MA, Dahlbom M, Huang LL, Marvi MM, Tang M, et al. Predicting seizure-free status for temporal lobe epilepsy patients undergoing surgery: prognostic value of quantifying maximal metabolic asymmetry extending over a specified proportion of the temporal lobe. *J Nucl Med.* (2007) 48:776–82. doi: 10.2967/jnumed.106.034249
28. Jeong SW, Lee SK, Kim KK, Kim H, Kim JY, Chung CK. Prognostic factors in anterior temporal lobe resections for mesial temporal lobe epilepsy: multivariate analysis. *Epilepsia.* (1999) 40:1735–9. doi: 10.1111/j.1528-1157.1999.tb01591.x
29. Hennessy MJ, Elwes RD, Binnie CD, Polkey CE. Failed surgery for epilepsy. A study of persistence and recurrence of seizures following temporal resection. *Brain.* (2000) 123:2445–66. doi: 10.1093/brain/123.12.2445
30. Lopez-Gonzalez MA, Gonzalez-Martinez JA, Jehi L, Kotagal P, Warbel A, Bingaman W. Epilepsy surgery of the temporal lobe in pediatric population: a retrospective analysis. *Neurosurgery.* (2012) 70:684–92. doi: 10.1227/NEU.0b013e318235183d
31. Elwan SA, So NK, Enatsu R, Bingaman WE. Pseudotemporal ictal patterns compared with mesial and neocortical temporal ictal patterns. *J Clin Neurophysiol.* (2013) 30:238–46. doi: 10.1097/WNP.0b013e3182872f70
32. LoPinto-Khoury C, Sperling MR, Skidmore C, Nei M, Evans J, Sharan A, et al. Surgical outcome in PET-positive, MRI-negative patients with temporal lobe epilepsy. *Epilepsia.* (2012) 53:342–8. doi: 10.1111/j.1528-1167.2011.03359.x
33. Bouter C, Bouter Y. (18)F-FDG-PET in mouse models of Alzheimer's disease. *Front Med.* (2019) 6:71. doi: 10.3389/fmed.2019.00071
34. Barba C, Barbati G, Minotti L, Hoffmann D, Kahane P. Ictal clinical and scalp-EEG findings differentiating temporal lobe epilepsies from temporal 'plus' epilepsies. *Brain.* (2007) 130:1957–67. doi: 10.1093/brain/awm108
35. Franke TN, Irwin C, Bayer TA, Brenner W, Beindorff N, Bouter C, et al. *In vivo* imaging with (18)F-FDG- and (18)F-florbetaben-PET/MRI detects pathological changes in the brain of the commonly used 5XFAD mouse model of Alzheimer's disease. *Front Med.* (2020) 7:529. doi: 10.3389/fmed.2020.00529
36. Ryvlin P, Cross JH, Rheims S. Epilepsy surgery in children and adults. *Lancet Neurol.* (2014) 13:1114–26. doi: 10.1016/S1474-4422(14)70156-5
37. Bartolomei F, Lagarde S, Wendling F, McGonigal A, Jirsa V, Guye M, et al. Defining epileptogenic networks: contribution of SEEG and signal analysis. *Epilepsia.* (2017) 58:1131–47. doi: 10.1111/epi.13791
38. Spencer DD, Gerrard JL, Zaveri HP. The roles of surgery and technology in understanding focal epilepsy and its comorbidities. *Lancet Neurol.* (2018) 17:373–82. doi: 10.1016/S1474-4422(18)30031-0
39. Chen D, Lu J, Zhou H, Jiang J, Wu P, Guo Q, et al. Glucose metabolic brain network differences between Chinese patients with lewy body dementia and healthy control. *Behav Neurol.* (2018) 2018:8420658. doi: 10.1155/2018/8420658
40. Andrews JP, Gummadavelli A, Farooque P, Bonito J, Arencibia C, Blumenfeld H, et al. Association of seizure spread with surgical failure in epilepsy. *JAMA Neurol.* (2018) 76:462−9. doi: 10.1001/jamaneurol.2018.4316
41. Bernhardt BC, Bernasconi A, Liu M, Hong SJ, Caldairou B, Goubran M, et al. The spectrum of structural and functional imaging abnormalities in temporal lobe epilepsy. *Ann Neurol.* (2016) 80:142–53. doi: 10.1002/ana.24691

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 23 December 2020
doi: 10.3389/fmed.2020.616746

Development of Combination Methods for Detecting Malignant Uptakes Based on Physiological Uptake Detection Using Object Detection With PET-CT MIP Images

Masashi Kawakami[1], *Kenji Hirata*[2], *Sho Furuya*[2], *Kentaro Kobayashi*[2], *Hiroyuki Sugimori*[3*], *Keiichi Magota*[4] *and Chietsugu Katoh*[1,3*]

[1] Graduate School of Biomedical Science and Engineering, Hokkaido University, Sapporo, Japan, [2] Department of Diagnostic Imaging, Graduate School of Medicine, Hokkaido University, Sapporo, Japan, [3] Faculty of Health Sciences, Hokkaido University, Sapporo, Japan, [4] Division of Medical Imaging and Technology, Hokkaido University Hospital, Sapporo, Japan

OPEN ACCESS

Edited by:
Kuangyu Shi,
University of Bern, Switzerland

Reviewed by:
Maria Vittoria Mattoli,
University of Studies G. d'Annunzio Chieti and Pescara, Italy
Bijoy Kundu,
University of Virginia, United States

***Correspondence:**
Hiroyuki Sugimori
sugimori@hs.hokudai.ac.jp
Chietsugu Katoh
chtgkato@med.hokudai.ac.jp

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: *13 October 2020*
Accepted: *04 December 2020*
Published: *23 December 2020*

Citation:
Kawakami M, Hirata K, Furuya S, Kobayashi K, Sugimori H, Magota K and Katoh C (2020) Development of Combination Methods for Detecting Malignant Uptakes Based on Physiological Uptake Detection Using Object Detection With PET-CT MIP Images. Front. Med. 7:616746. doi: 10.3389/fmed.2020.616746

Deep learning technology is now used for medical imaging. YOLOv2 is an object detection model using deep learning. Here, we applied YOLOv2 to FDG-PET images to detect the physiological uptake on the images. We also investigated the detection precision of abnormal uptake by a combined technique with YOLOv2. Using 3,500 maximum intensity projection (MIP) images of 500 cases of whole-body FDG-PET examinations, we manually drew rectangular regions of interest with the size of each physiological uptake to create a dataset. Using YOLOv2, we performed image training as transfer learning by initial weight. We evaluated YOLOv2's physiological uptake detection by determining the intersection over union (IoU), average precision (AP), mean average precision (mAP), and frames per second (FPS). We also developed a combination method for detecting abnormal uptake by subtracting the YOLOv2-detected physiological uptake. We calculated the coverage rate, false-positive rate, and false-negative rate by comparing the combination method-generated color map with the abnormal findings identified by experienced radiologists. The APs for physiological uptakes were: brain, 0.993; liver, 0.913; and bladder, 0.879. The mAP was 0.831 for all classes with the IoU threshold value 0.5. Each subset's average FPS was 31.60 ± 4.66. The combination method's coverage rate, false-positive rate, and false-negative rate for detecting abnormal uptake were 0.9205 ± 0.0312, 0.3704 ± 0.0213, and 0.1000 ± 0.0774, respectively. The physiological uptake of FDG-PET on MIP images was quickly and precisely detected using YOLOv2. The combination method, which can be utilized the characteristics of the detector by YOLOv2, detected the radiologist-identified abnormalities with a high coverage rate. The detectability and fast response would thus be useful as a diagnostic tool.

Keywords: object detection, deep learning, positron emission tomography, YOLOv2, computer vision

INTRODUCTION

Deep learning technology has developed rapidly and is now used in real-world settings such as automated driving, games, image processing, and voice recognition (1–4). Deep learning has also been applied to the field of medical imaging; e.g., in the classification of computed tomography (CT) images in different slice positions (5) and its training algorithm (6), research concerning the diagnosis and processing of pulmonary nodules by deep learning for feature extraction, detection, false-positive reduction, and benign malignant classification (7), and a study using deep learning to improve the performance of the automatic detection of lesions on mammograms (8). In such applications, knowledge of anatomy is required to process and diagnose medical images, and an inexperienced person may not be able to process and diagnose the images appropriately. However, if automatic object detection using deep learning (9–14) can be used for medical imaging, it could be possible to perform highly reproducible processing without the requirement of the knowledge and experience of physicians and radiologists.

Fluorodeoxyglucose-positron emission tomography (FDG-PET) (15) is an imaging method in which FDG labeled with fluorine-18 (F-18) is injected into the body, and two 511 keV annihilation photons which are produced by the positron decay of F-18 are simultaneously injected into the opposing detectors and reconstructed. FDG is an analog of glucose and accumulates in tumors with increased glucose metabolism as well as in organs *in vivo*, such as the brain, where glucose consumption as energy is high. It is therefore necessary to determine whether each site of high FDG uptake is a physiological uptake or an abnormal uptake. It has been demonstrated that a convolutional neural network (CNN) was useful for classifying FDG-PET images into normal, abnormal, and equivocal uptakes (16). In the present study, we investigated the precision of an object detection model, You Only Look Once version 2 (YOLOv2) (17), which uses deep learning to automatically detect the physiological uptakes on maximum intensity projection (MIP) images of FDG-PET in a rectangular region. We also developed a combination method to generate images in which abnormal uptakes were enhanced by subtracting only the detected physiological uptakes from the original MIP images. For an evaluation of the potential clinical uses of this combination method, we calculated the coverage rate by comparing the generated images to the abnormal uptakes that were identified by previous imaging findings.

MATERIALS AND METHODS

Subject and PET-CT Scans

The study included a total of 500 patients (287 males and 213 females, age 61.3 ± 17.0 years [mean ± SD]) who underwent a whole-body FDG-PET examination for the screening of malignant tumors between January and May 2016 at our institute. All MIP images were acquired using either a GEMINI TF64 PET-CT scanner (Philips Healthcare, Cleveland, OH, USA), or a Biograph64 PET-CT scanner (Siemens Healthcare, Erlangen, Germany). This study was approved by our institute's Ethics Committee [#017-0365].

Automatic Detection

The Creation of the Datasets

A total of 3,500 MIP images of the 500 patients (seven images per patient at every 10° to ±30° from the front) were generated in the workstation equipped with the PET-CT scanners. We defined five classes of physiological uptake to be automatically detected: brain, heart, liver, kidney, and bladder. All image data were converted from Digital Imaging and Communications in Medicine (DICOM) files to Joint Photographic Experts Group (JPEG) files for further use.

The JPEG files were loaded into the in-house MATLAB software program (MATLAB2019b, The MathWorks, Natick, MA, USA), and this program was used to draw rectangular regions of interest (ROIs) to enclose each physiological uptake (**Figure 1**). The ROI data were outputted as a text file, which included the object name, the coordinates, and the size of each ROI. We divided the supervised data into five subsets for nested cross-validation (18). Each 100 patients contributed 700 images; we used 2,800 images from 400 patients for training, and the remaining 700 images for testing (**Figure 2**). Each subset was an independent combination of 400 patients for training and 100 patients for test images, to prevent the mixing of patient images between the training and testing images within the subsets. To effectively learn for the training dataset, we performed data augmentation (19, 20) using image rotation from −15° to 15° in 3° steps and a zoom rate from 0.9 to 1.1 in 0.1 steps.

Training Images for Model Creation

We developed a software program for object detection with a deep learning technique via the in-house MATLAB software; we used a deep learning-optimized machine with an Nvidia Quadro P5000 graphics card (Nvidia Corp., Santa Clara, CA), which provides 8.9 Tera floating-point single-precision operations per sec, 288 GB/sec memory bandwidth, and 16 GB memory per board. We performed the image training as transfer learning by initial weight using YOLOv2, with the MATLAB deep learning Toolbox and Computer Vision System Toolbox. The training model hyperparameters were as follows: maximum training epochs, 10; initial learning rate, 0.00001; mini-batch size, 96. We used stochastic gradient descent with momentum for optimization with an initial learning rate. We set the momentum and L2 regulation to 0.9 and 0.0001, respectively. We performed image training five times based on the training subsets shown in **Figure 2**.

Evaluation of the Created Models

We incorporated the predicted bounding boxes into the MATLAB software in order to reveal the region of each physiological uptake as a bounding box. We also evaluated each physiological uptake by determining the average precision (AP), the mean average precision (mAP) (21), and the frames per second (FPS) for an estimation of the efficiency of the created model. The AP and mAP values were calculated by each intersection over union (IoU). We examined the bounding boxes based on the supervised ROI according to the "evaluateDetectionPrecision" and "evaluateDetectionMissRate" functions in the MATLAB Computer Vision Toolbox.

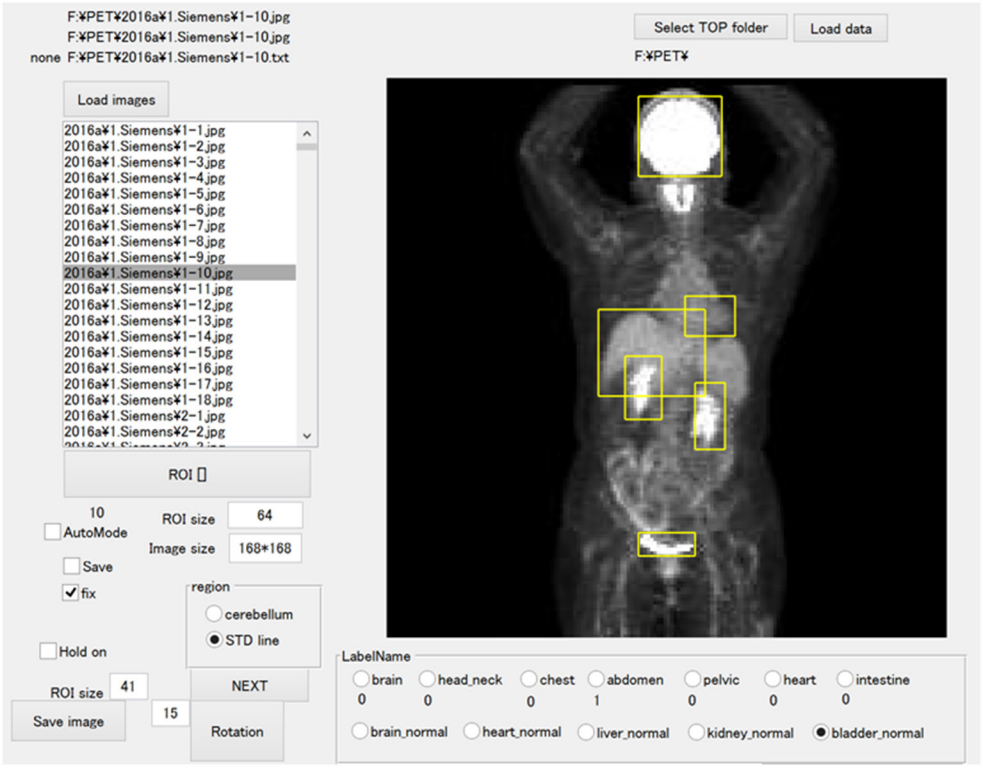


FIGURE 1 | The software for training outlined ROIs over the physiological uptakes. The yellow bounding boxes enclose physiological uptakes.

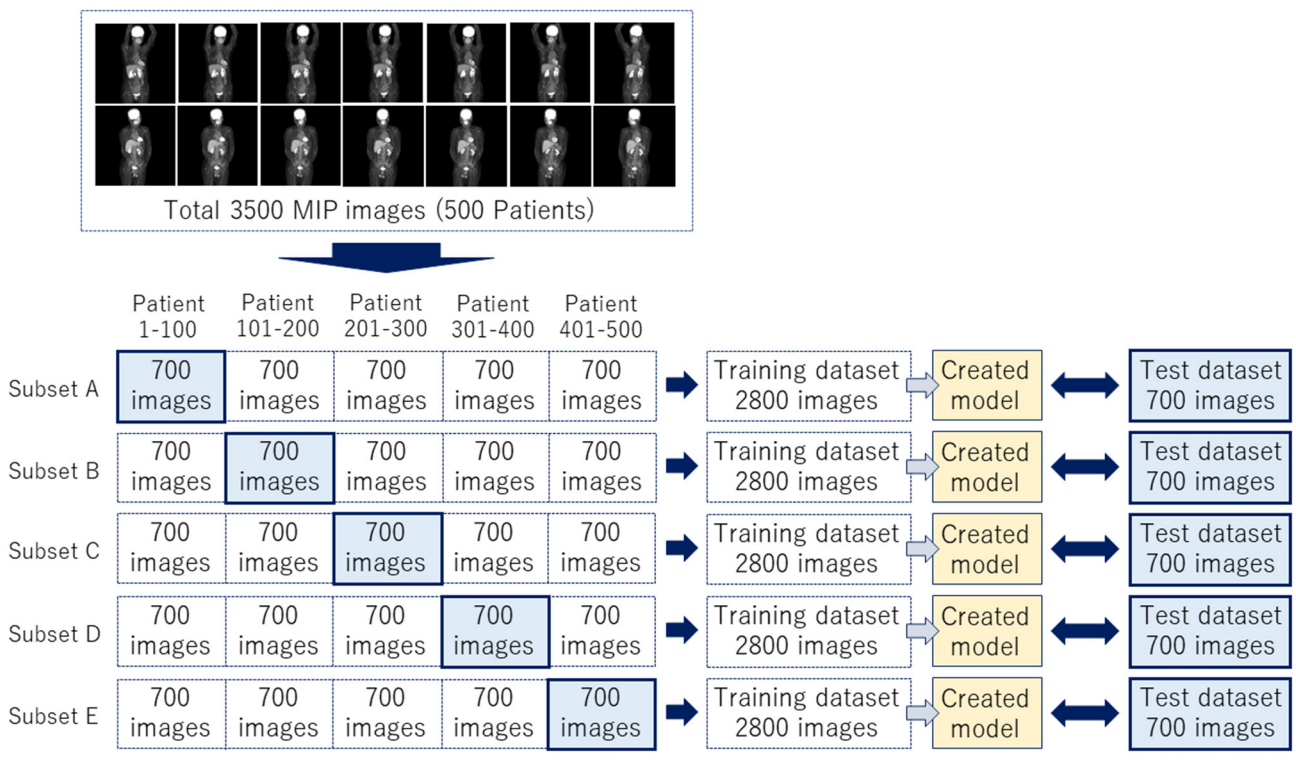


FIGURE 2 | A total of 3,500 images were divided into five subsets to complete a nested cross-validation.

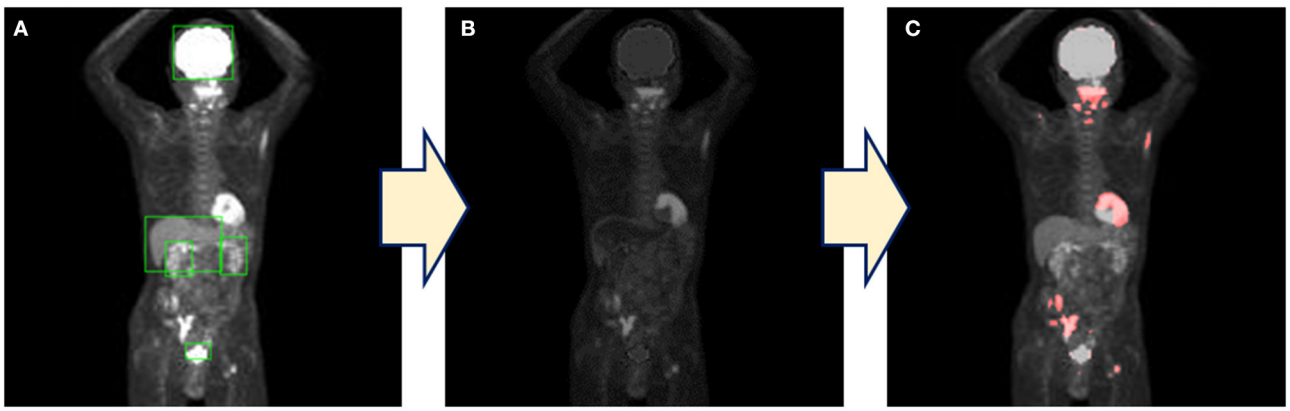


FIGURE 3 | The process of the combination method. **(A)** The detection of physiological uptake by YOLOv2. **(B)** Only physiological uptakes detected by YOLOv2 were subtracted from the original images. **(C)** A color map is generated by coloring the pixels above the threshold.

Combination Method

The Creation of Color Maps

Each physiological uptake (brain, heart, liver, kidney, and bladder) detected by the created model described above in section **Automatic Detection** (**Figure 3A**) was cropped from the original MIP image with a five-pixel margin on the coordinate information of the bounding boxes. For the clipped physiological uptake images, we performed threshold processing to set <50% of the maximum pixel values to zero, and we applied a two-dimensional (2D) Gaussian smoothing kernel with a standard deviation of 1.2. The images generated by this process were subtracted from the original MIP images (**Figure 3B**); the histogram of the images was drawn, and the mode of frequency was determined. The threshold was defined as the value of the pixel value plus 100. Higher uptakes other than physiological uptakes were emphasized in the image by displaying red pixels above the threshold value (**Figure 3C**). Our new combination method was thus defined as the generation of these images as a color map derived from the series of methods described above.

Evaluation of the Created Color Maps

We evaluated the combination method by comparing the abnormal imaging findings between the abnormal findings obtained by two experienced radiologists (SF, 5 years; KH, 18 years) and the color maps generated by the combination method. The radiologist's findings were evaluated according to the presence/absence of abnormalities in each of the seven regions (brain, head/neck, chest, abdomen, pelvis, heart, intestine). When the region colored on the color map corresponded to the region diagnosed as abnormal by the radiologist, it was considered to be correctly detected.

We defined the coverage rate as the percentage of correctly detected abnormalities relative to the radiologist's findings of the presence of abnormalities. A false-positive result was defined as when a site with no abnormalities on the radiologist's findings was colored on the color map. And also, a false-negative result was defined as when the site noted as abnormal by the radiologist was not colored on the color map. We defined the false-positive rate as the ratio of false-positive results to the radiologist's findings of no abnormalities and the false-negative rate as the ratio of false-negative results to the site of colored on the color map. In addition, we obtained false-positive and false-negative rates for each site. These values were calculated for the evaluation of the detection precision of the combination method.

RESULTS

Average Precision of Each Class

The average precision of each class automatically detected by YOLOv2 is shown in **Figure 4**. At the IoU of 0.5, physiological uptakes in the brain were detected with rather high precision (AP: 0.993), followed by high APs in the liver (0.913), bladder (0.879), and kidneys (0.843). The detection of the cardiac uptakes were the worst, with an AP of 0.527 at the IoU of 0.5.

Figure 5 shows the mean average precision of each physiological uptake (brain, heart, liver, kidney, and bladder) detected by YOLOv2. The mAP value was decreased over the threshold IoU of 0.5. The mAP was 0.831 with the threshold IoU of 0.5. The average FPS for each subset was 31.60 ± 4.66.

Coverage Rate, False Positive Rate, and False Negative Rate of the Combination Method

Table 1 shows the coverage and false-positive rate of the color maps generated by the combination method. The coverage rate, false-positive rate, and false-negative rate for detecting abnormal uptake were 0.9205 ± 0.0312, 0.3704 ± 0.0213, and 0.1000 ± 0.0774, respectively.

Table 2 shows the false-positive rate and false-negative rate by site by the combination method. The false-positive rate was highest in the head/neck, at 0.7629 ± 0.0385. The false-negative rates were higher in the abdomen and head/neck, 0.2047 ± 0.1254 and 0.2000 ± 0.4472, respectively.

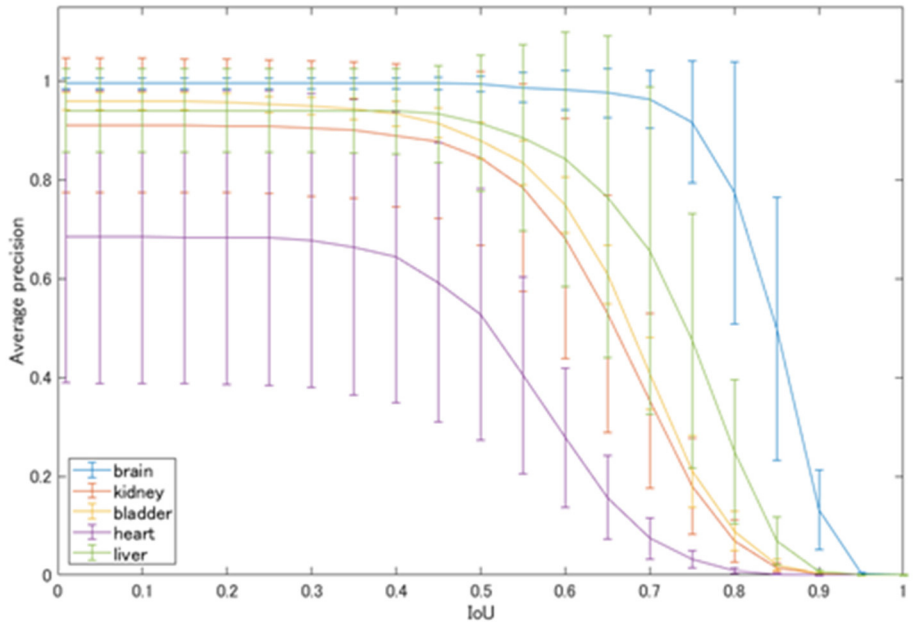


FIGURE 4 | Average precision of each physiological uptake.

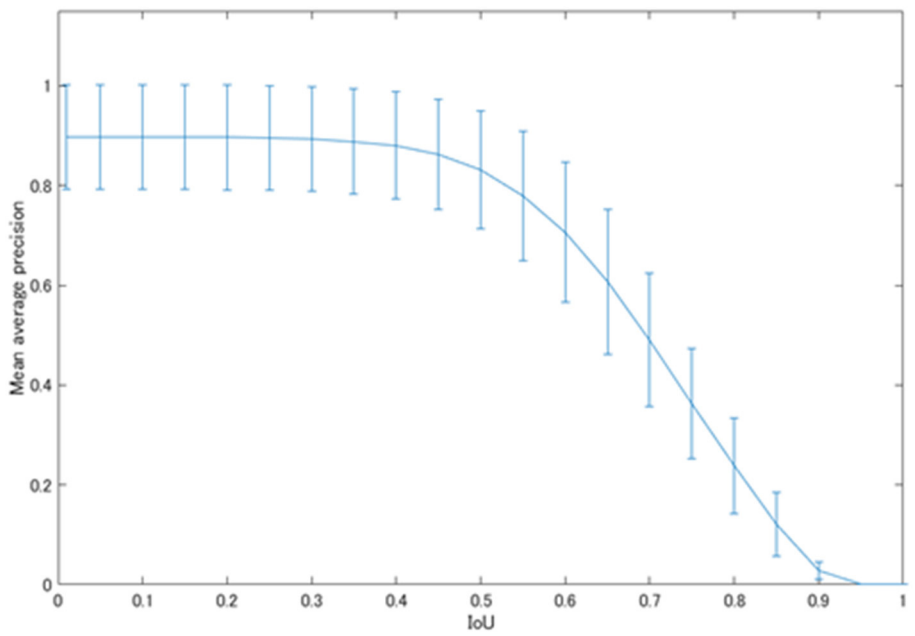


FIGURE 5 | The mAP values of the physiological uptakes.

DISCUSSION

We evaluated the precision of a network model of deep learning for object detection (i.e., YOLOv2) for detecting physiological uptakes on FDG-PET images with a rectangular region, and we developed a combination method using YOLOv2 and subtraction processing for the detection of abnormal uptakes. The detector was created by training with a dataset of 3,500 MIP images with data augmentation processes (such as rotation and zooming) and the mAP of 0.831 with the IoU of 0.5. The average FPS was >30. The results demonstrated high detection precision and a high speed for the detection of physiological uptakes. In particular, the

TABLE 1 | Coverage rate, false-positive rate, and false-negative rate of the new combination method.

	Coverage rate	False positive rate	False negative rate
Subset A	0.8719	0.3796	0.2310
Subset B	0.9343	0.3356	0.0543
Subset C	0.9554	0.3684	0.0329
Subset D	0.9129	0.3857	0.0970
Subset E	0.9279	0.3828	0.0851
Mean ± SD	0.9205 ± 0.0312	0.3704 ± 0.0213	0.1000 ± 0.0774

TABLE 2 | False-positive rate and false-negative rate by site.

	False positive rate	False negative rate
Brain	0.0137 ± 0.0254	0.2000 ± 0.4472
Head/neck	0.7629 ± 0.0385	0
Chest	0.3523 ± 0.1460	0.0331 ± 0.0197
Abdomen	0.3326 ± 0.1873	0.2047 ± 0.1254
Pelvis	0.1674 ± 0.0341	0.0667 ± 0.0726
Heart	0.0331 ± 0.0296	0.0761 ± 0.0288
Intestine	0.3754 ± 0.0602	0.1198 ± 0.1539

physiological uptakes of the brain could be detected with fairly high precision, with an AP of 0.993 with the IoU of 0.5. The next-highest values were found for the liver, bladder and kidneys at 0.913, 0.879, and 0.843, respectively.

However, as shown in **Figure 4**, the detection of physiological uptakes in the heart resulted in lower APs compared to those in the other classes. One possible reason for the very high APs (>0.9) in physiological uptakes in the brain and liver compared to the slightly lower APs in other classes is that FDG-PET provides metabolic images rather than anatomical images like CT or magnetic resonance imaging (MRI). The brain and liver showed no differences in the MIP images due to the minimal differences in metabolism between individuals. However, the shape and degree of uptakes of the bladder and kidneys varied greatly on the images, as there were large individual differences in uptakes depending on the degree of urination. In addition, normal heart uptakes were more difficult to diagnose than the other four classes because of the variety of uptake patterns (22), which may have contributed to the low AP for the heart.

With regard to limitations of the detection of physiological uptakes, the number of features for the detector that was necessary for the training of the created models was limited. In other words, for the further improvement of the detection precision, we have to consider increasing the number of training images and the number of patterns of data augmentation because the present study was performed with rotation and zooming of the images. However, our findings demonstrated that higher AP and mAP values could be obtained by varying the degree of rotation and the zoom rate as data augmentation. In addition, regarding the pixel data, the raw data of the DICOM files (which had the dynamic range of pixel values) would be taken into account for the evaluation of the new method's precision because the present study was performed using JPEG images. Moreover, it has been reported that the mAP was improved by color operations and geometric operations (23), and the precision of the mAP could be changed by performing procedures other than those used in the present study.

Regarding the CNN models, although our results showed sufficient precision and response speed for the real-time detection, further improvements of the detection precision and speed may be obtained by using a network model for object detection other than YOLOv2, such as DetectNet (24), Single Shot MultiBox Detector (25), and Faster R-CNN (26). New network models such as Feature Pyramid Networks (27) and Mask R-CNN (28), which are based on Faster R-CNN with additional technologies, have been reported to improve processing speed and average precision. There is also an improved version of YOLOv2 with a deeper network model, YOLOv3 (29). However, the detection precision shown as the mAP was 0.831 with the IoU of 0.5, and the average FPS was over 30 FPS in the present study. These results demonstrated that the detection was sufficiently accurate and faster compared to other studies aimed at real-time detection (30). Therefore, although there is room for further improvement in the detection performance due to factors such as the number of training images, data augmentation, and different network models, we observed that the detection performance obtained herein (including the speed response) was sufficient for use in image diagnoses.

We took advantage of the detection of physiological uptake by the detector created using YOLOv2, an object detection model based on deep learning technology. The combination method was our newly developed method for detecting abnormal uptakes by combining YOLOv2 and a subtraction process. The color maps generated by the combination method showed great merit, with a very high coverage rate of >92% for the abnormal findings identified by the highly experienced radiologist. We used MIP data in this study because of the smaller size of the data and the ability to evaluate the whole body in a single image. The use of MIP images not only reduced the learning time even when the number of images was increased but also provides more information in one image than can be obtained when using tomographic images. For example, because it is often difficult to distinguish between normal and abnormal uptakes in a single tomographic image, the diagnosis is usually made by examining both the upper and lower slices. For these reasons, there are advantages to using MIP images for diagnoses using deep learning.

However, the false-positive rate obtained by this detector was ∼37%, which is not very low; this result might be due to the use of MIP images with 2D data. The generation of a false high uptake might be caused by overlapping of low uptake when the three-dimensional (3D) structure of the body was rendered into a 2D image, and this effect might be the reason for the increase of false-positive results. It was thus difficult to reduce the number of false positives with MIP images, and there was a limit to the precision of the combination method with MIP

data. Incorporating tomographic data could be one solution. Moreover, the false-positive rate by site was the highest in the head/neck. As for the neck region, Purohit BS reported (31) that FDG uptake by normal lymphoid tissue act as a confounding factor for the diagnosis of neck tumors. This was the reason we did not define the uptakes at this site as a physiological uptake. As a result, the false-positive rate was higher than any other site due to detecting physiological uptakes incorrectly. However, the head/neck was the only site with a false-negative rate of zero and did not miss the lesion. Regarding to the false-negative rate, it was higher in the abdomen and brain because abnormal uptakes within physiological uptakes were more common in these sites than in other sites, and these uptakes could not be detected.

Although we applied the dataset to the 2D network model as YOLOv2 due to the limitation of computer resources, there are 3D network models (32–36) that can be used with one-time training with the whole data. We have also considered these 3D network models for the evaluation of precision in a future study. Furthermore, we have to take into consideration changing window width of the MIP images to distinguish between physiological uptakes and tumors because the MIP technique has a limitation to detect abnormal uptakes within physiological uptakes. However, in light of the accurate detection of the physiological uptakes observed herein and the coverage rate for abnormal uptake indicated by a radiologist, our present results have established this combination method as a useful diagnostic tool with real-time detection.

CONCLUSIONS

We investigated the precision of the detection of physiological uptakes and developed a combination method for diagnoses based on FDG-PET images. With the use of using YOLOv2, the physiological uptake of FDG on MIP images was automatically detected with high precision and high speed. In addition, the combination method, which utilizes the characteristics of the detector by YOLOv2, detected abnormalities identified by the experienced radiologist with a high coverage rate. The combination method's detection performance and fast response demonstrated its usefulness as a diagnostic aid tool.

DATA AVAILABILITY STATEMENT

The original contributions generated for this study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Hokkaido University Hospital Ethics Committee. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MK contributed to data analysis and the writing and editing of the article. KH, SF, and KK proposed the idea and reviewed and edited the paper. HS contributed to algorithm construction, performed supervision, and reviewing and editing. KM contributed to data acquisition and reviewed and edited the paper. CK performed supervision and project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was partly supported by a JSPS KAKENHI grant, no. JP19K17127.

REFERENCES

1. Jana S, Tian Y, Pei K, Ray B. DeepTest: automated testing of deep-neural-network-driven autonomous cars. *Proc - Int Conf Softw Eng*. (2018) 303–14. doi: 10.1145/3180155.3180220
2. Mnih V, Silver D. Playing atari with deep reinforcement learning. *arXiv*. (2013) 1–9. Available online at: https://arxiv.org/abs/1312.5602 (accessed October 12, 2020).
3. Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: overview, challenges and future. *Classifi BioApps*. (2017) 26:1–30. doi: 10.1007/978-3-319-65981-7_12
4. Liptchinsky V, Synnaeve G, Collobert R. *Letter-Based Speech Recognition with Gated ConvNets*. (2017). p. 1–13. Available online at: http://arxiv.org/abs/1712.09444 (accessed October 12, 2020).
5. Sugimori H. Classification of computed tomography images in different slice positions using deep learning. *J Healthc Eng*. (2018) 2018:1753480. doi: 10.1155/2018/1753480
6. Sugimori H. Evaluating the overall accuracy of additional learning and automatic classification system for CT images. *Appl Sci*. (2019) 9:682. doi: 10.3390/app9040682
7. Yang Y, Feng X, Chi W, Li Z, Duan W, Liu H, et al. Deep learning aided decision support for pulmonary nodules diagnosing: a review. *J Thorac Dis*. (2018) 10:S867–75. doi: 10.21037/jtd.2018.0257
8. Suzuki S, Zhang X, Homma N, Yoshizawa M. Deep convolutional neural network detecting masses in mammograms based on transfer learning of a deep convolutional neural network. In: *Proceedings of 2016 Tohoku-Section Joint Convention of Institutes of Electrical and Information Engineers*. (2016). Available online at: https://www.jstage.jst.go.jp/article/tsjc/2016/0/2016_1/_pdf (accessed October 12, 2020).
9. Sugimori H, Kawakami M. Automatic detection of a standard line for brain magnetic resonance imaging using deep learning. *Appl Sci*. (2019) 9:3849. doi: 10.3390/app9183849
10. Szegedy C, Toshev A, Erhan D. Deep neural networks for object detection. *Adv Neural Inf Process Syst*. (2013) 260:2553–61. Available online at: https://papers.nips.cc/paper/5207-deep-neural-networks-for-object-detection.pdf (accessed October 12, 2020).
11. Cao Z, Duan L, Yang G, Yue T, Chen Q. An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. *BMC Med Imaging*. (2019) 19:1–9. doi: 10.1186/s12880-019-0349-x

12. König D. *Deep Learning for Person Detection in Multi-Spectral Videos*. Thesis (2017). Available online at: https://d-nb.info/1135265593/34 (accessed October 12, 2020).
13. Li M, Yu C, Nian F, Li X. A face detection algorithm based on haar-like t features. *Int J Hybrid Inf Technol.* (2015) 8:285–96. doi: 10.14257/ijhit.2015.8.1124
14. Rudiawan E, Analia R, DSP, Soebakti H. The deep learning development for real-time ball and goal detection of Barelang-FC. In: *2017 International Electronics Symposium on Engineering Technology and Applications (IES-ETA)*. Surabaya (2017). p. 146–51. doi: 10.1109/ELECSYM.2017.8240393
15. Jamar F, Buscombe J, Chiti A, Christian PE, Delbeke D, Donohoe KJ, et al. EANM/SNMMI guideline for 18F-FDG use in inflammation and infection. *J Nucl Med*. (2013) 54:647–58. doi: 10.2967/jnumed.112112524
16. Kawauchi K, Furuya S, Hirata K, Katoh C, Manabe O, Kobayashi K, et al. A convolutional neural network-based system to classify patients using FDG PET/CT examinations. *BMC Cancer*. (2020) 20:1–10. doi: 10.1186/s12885-020-6694-x
17. Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI (2017). p. 6517–25. doi: 10.1109/CVPR.2017.690
18. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS ONE*. (2019) 14:1–20. doi: 10.1371/journal.pone0224365
19. Zhang C, Zhou P, Li C, Liu L. A convolutional neural network for leaves recognition using data augmentation. *Proc - 15th IEEE Int Conf Comput Inf Technol CIT 2015, 14th IEEE Int Conf Ubiquitous Comput Commun IUCC 2015, 13th IEEE Int Conf Dependable, Auton Se*. (2015) 2143–50. doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.318
20. Hussain Z, Gimenez F, Yi D, Rubin D. Differential data augmentation techniques for medical imaging classification tasks. *AMIA Annu Symp proceedings AMIA Symp*. (2017) 2017:979–84. Available online at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977656/ (accessed October 12, 2020).
21. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. *Int J Comput Vis*. (2010) 88:303–38. doi: 10.1007/s11263-009-0275-4
22. Nose H, Otsuka H, Otomi Y, Terazawa K, Takao S, Iwamoto S, et al. The physiological uptake pattern of 18F-FDG in the left ventricular myocardium of patients without heart disease. *J Med Investig*. (2014) 61:53–8. doi: 10.2152/jmi.6153
23. Zoph B, Cubuk ED, Ghiasi G, Lin T-Y, Shlens J, Le QV. Learning *Data Augmentation Strategies for Object Detection*. (2019). Available online at: http://arxiv.org/abs/1906.11172 (accessed October 12, 2020).
24. Holupka EJ, Rossman J, Morancy T, Aronovitz J, Kaplan ID. The detection of implanted radioactive seeds on ultrasound images using convolution neural networks. In: *MIDL 2018 Conference*. (2018). p. 1–8. Available online at: https://openreview.net/forum?id=H1hlwg2oG (accessed October 12, 2020).
25. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. SSD: single shot multibox detector. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. (2016) 9905:21–37. doi: 10.1007/978-3-319-46448-0_2
26. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. (2017) 39:1137–49. doi: 10.1109/TPAMI.20162577031
27. Lin T, Doll P, Girshick R, He K, Hariharan B, Belongie S, et al. Feature pyramid networks for object detection. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2017). p. 936–44. doi: 10.1109/CVPR.2017106
28. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell*. (2020) 42:386–97. doi: 10.1109/TPAMI.20182844175
29. Redmon J, Farhadi A. *YOLOv3: An Incremental Improvement*. (2018). Available online at: http://arxiv.org/abs/1804.02767 (accessed October 12, 2020).
30. Sugimori H, Sugiyama T, Nakayama N, Yamashita A, Ogasawara K. Development of a deep learning-based algorithm to detect the distal end of a surgical instrument. *Appl Sci*. (2020) 10:4245. doi: 10.3390/app10124245
31. Purohit BS, Ailianou A, Dulguerov N, Becker CD, Ratib O, Becker M. FDG-PET/CT pitfalls in oncological head and neck imaging. *Insights Imaging*. (2014) 5:585–602. doi: 10.1007/s13244-014-0349-x
32. Qi CR, Nießner M, Dai A, Guibas LJ. Volumetric and multi-view CNNs for object classification on 3D data supplementary material. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV (2016). p. 5648–56. doi: 10.1109/CVPR.2016.609
33. Maturana D, Scherer S. VoxNet: a 3D convolutional neural network for real-time object recognition. *IEEE Int Conf Intell Robot Syst*. (2015) 2015:922–8. doi: 10.1109/IROS.20157353481
34. CR Qi, Liu W, Wu C, Su H, Guibas LJ. Frustum pointnets for 3D object detection from RGB-D data. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. (2018) 918–27. doi: 10.1109/CVPR.201800102
35. Su H, Maji S, Kalogerakis E, Learned-Miller E. Multi-view convolutional neural networks for 3D shape recognition. *Proc IEEE Int Conf Comput Vis*. (2015) 2015:945–53. doi: 10.1109/ICCV.2015114
36. Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, et al. 3D ShapeNets: a deep representation for volumetric shapes. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. (2015) 1912–20. doi: 10.1109/CVPR.2015.7298801

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 23 December 2020
doi: 10.3389/fmed.2020.603303

Texture Analysis in the Diagnosis of Primary Breast Cancer: Comparison of High-Resolution Dedicated Breast Positron Emission Tomography (dbPET) and Whole-Body PET/CT

Yoko Satoh[1,2]*, *Kenji Hirata*[3], *Daiki Tamada*[2], *Satoshi Funayama*[2] and *Hiroshi Onishi*[2]

[1] Yamanashi PET Imaging Clinic, Yamanashi, Japan, [2] Department of Radiology, University of Yamanashi, Yamanashi, Japan, [3] Department of Diagnostic Imaging, School of Medicine, Hokkaido University, Sapporo, Japan

OPEN ACCESS

Edited by:
Xiu Ying Wang,
The University of Sydney, Australia

Reviewed by:
Gaurav Malviya,
University of Glasgow,
United Kingdom
Linlin Wang,
Shandong University, China
Hui Cui,
La Trobe University, Australia

***Correspondence:**
Yoko Satoh
ysatoh@ypic.jp

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: 06 September 2020
Accepted: 02 December 2020
Published: 23 December 2020

Citation:
Satoh Y, Hirata K, Tamada D, Funayama S and Onishi H (2020) Texture Analysis in the Diagnosis of Primary Breast Cancer: Comparison of High-Resolution Dedicated Breast Positron Emission Tomography (dbPET) and Whole-Body PET/CT. Front. Med. 7:603303. doi: 10.3389/fmed.2020.603303

Objective: This retrospective study aimed to compare the ability to classify tumor characteristics of breast cancer (BC) of positron emission tomography (PET)-derived texture features between dedicated breast PET (dbPET) and whole-body PET/computed tomography (CT).

Methods: Forty-four BCs scanned by both high-resolution ring-shaped dbPET and whole-body PET/CT were analyzed. The primary BC was extracted with a standardized uptake value (SUV) threshold segmentation method. On both dbPET and PET/CT images, 38 texture features were computed; their ability to classify tumor characteristics such as tumor (T)-category, lymph node (N)-category, molecular subtype, and Ki67 levels was compared. The texture features were evaluated using univariate and multivariate analyses following principal component analysis (PCA). AUC values were used to evaluate the diagnostic power of the computed texture features to classify BC characteristics.

Results: Some texture features of dbPET and PET/CT were different between Tis-1 and T2-4 and between Luminal A and other groups, respectively. No association with texture features was found in the N-category or Ki67 level. In contrast, receiver-operating characteristic analysis using texture features' principal components showed that the AUC for classification of any BC characteristics were equally good for both dbPET and whole-body PET/CT.

Conclusions: PET-based texture analysis of dbPET and whole-body PET/CT may have equally good classification power for BC.

Keywords: dedicated breast positron emission tomography (dbPET), positron emission tomography/computed tomography (PET/CT), texture analysis, breast cancer, 18F-FDG

INTRODUCTION

Due to the recent advances in BC treatment, neo-adjuvant systemic chemotherapy is often performed before surgery. Therefore, highly accurate staging before treatment is essential. However, because BC is characterized by heterogeneity, it is difficult to predict tumor characteristics and prognosis from small specimens biopsied from a limited number of lesions. 18F-fluorodeoxyglucose (FDG) positron emission tomography (PET) can assess metabolic information on various tumors—a difficult task with conventional imaging modalities—and evaluate a wide range of pathological conditions in a minimally invasive manner. For this reason, PET is now widely used for benign vs. malignant lesion differentiation, staging, recurrence diagnosis, and prediction of prognosis in various types of cancer, including BC (1).

Until recently, the usefulness of local evaluation by whole-body PET/computed tomography (CT) has been limited by its limited spatial resolution and the physiological background accumulation in mammary glands (2). However, in more recent years, the performance of whole-body PET/CT scanners has increased along with their diagnostic ability in local evaluation due to the widespread use of devices using time-of-flight (TOF) or point-spread-function (PSF) methods for image reconstruction (3). Furthermore, high-resolution breast PET scanners have been developed to detect BC lesions smaller than those detectable by whole-body PET (4–6). By using the two devices in combination, it has been possible to accurately and more rapidly evaluate local and metastatic BC lesions.

In the past few years, research on radiomics has focused particularly on texture analysis using various imaging modalities, including BC studies utilizing MRI and US (7, 8). Some studies reported that PET images' texture features are associated with BC subtypes and prognosis (9–11). Texture analysis, which assesses intra-tumoral heterogeneity to compute image-specific information, is highly reproducible, has little variation among diagnostic radiologists, and will help mitigate the shortage of said radiologists who are excellent at diagnosing BC imaging. Its usefulness suggests that PET may contribute more widely to the diagnosis, treatment, and post-treatment management of BC.

Regarding the comparison between dbPET and whole-body PET/CT, there have been reports on the evaluation of the standard performance of the scanners and detectability of BC (12, 13), but none on the comparison of their diagnostic ability using texture analysis. This study aimed to compare the classifying ability of PET-derived texture features for BC's tumor characteristics between dbPET and whole-body PET/CT.

Abbreviations: AUC, area under the curve; AJCC, American Joint Committee on Cancer; BC, breast cancer; CT, computed tomography; dbPET, dedicated breast tomography PET; ER, estrogen receptor; FDG, fluorodeoxyglucose; FOV, field-of-view; GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run length matrix; GLZSM, gray-level zone size matrix; HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; ISH, *in situ* hybridization; MTV, metabolic tumor volume; N, lymph node staging; NGLDM, neighborhood gray-level difference matrix; PCA, principal component analysis; PET, positron emission tomography; PR, progesterone receptor; PSF, point-spread-function; ROC, receiver-operating characteristic; SUV, standardized uptake value; TLG, total lesion glycolysis; T, tumor staging; VOI, volume of interest.

MATERIALS AND METHODS

This single-institute, retrospective study was approved by the Institutional Review Board and the Ethics Committee of our institute and was carried out in accordance with the Declaration of Helsinki. Written informed consent for future access and anonymous use of their data was obtained from each patient.

Patients

We enrolled 798 consecutive women who underwent dbPET and whole-body PET/CT at our institute between April 2015 and March 2018. BCs that were selected fulfilled the following inclusion criteria: (1) the molecular subtype of BC had been determined; (2) patient clinical history was available. The exclusion criteria were as follows: (1) patients with a history of other malignancies; (2) missing or incomplete clinical data; (3) patients undergoing chemotherapy or within 1 year after its completion; (4) BC not successfully extracted in both dbPET and PET/CT because the SUV of the background mammary gland was higher than 40% of the BC SUVmax.

Subtype Classification

BC diagnoses through tumor histology and immunohistochemistry were made using surgical or biopsy specimens of core needle biopsy before neoadjuvant chemotherapy. Tumor, lymph nodes, and metastasis (TMN) categorization of malignant tumors was established following the 8th edition of the American Joint Committee on Cancer (AJCC) staging system (14). The molecular markers examined included estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor 2 amplified (HER2) expression. ER and PR status was considered positive for tumors showing at least 10% of positive cells. HER2 status was assessed by immunohistochemistry (IHC): Tumors scored as 3+ were classified as HER2 positive; those scored as 0 or 1+ were classified as HER2 negative. For tumors scored as 2+, further confirmation using molecular tests [*in situ* hybridization (ISH)] was obtained. ISH non-amplified tumors were classified as HER2 negative, and ISH amplified ones as HER2 positive. IHC classification followed the 13th St. Gallen International Breast Cancer Conference (2013) recommendations, with a Ki67 threshold of 20% (15). BCs were classified into four subtypes: (1) luminal A: ER and/or PR positive, HER2 negative, and low expression of Ki67 (<20%); (2) luminal B: (a) ER and/or PR positive, HER2 negative and high expression of Ki67 (20%≤), or (b) ER and/or PR positive, HER2 positive; (3) HER2: ER and PR negative, and HER2 positive; and (4) triple-negative: ER, PR, and HER2 negative.

Ring-Shaped dbPET Scanner

The ring-shaped dbPET scanner (Elmammo, Shimadzu Corporation, Kyoto, Japan) comprises a total of 36 detector modules (12 per ring) arranged in three continuous rings, has a diameter of 195 mm and axial length of 156.5 mm, and has depth-of-interaction measurement capability (16). The transaxial effective field-of-view (FOV) is 185 × 156.5 mm^2. Each detector block consists of a four-layered 32 × 32 array of lutetium

oxyorthosilicate crystals coupled to a 64-channel positron-sensitive photomultiplier tube via a light guide. Attenuation correction was calculated using a uniform attenuation map with object boundaries obtained from emission data (17). Scatter correction was performed using the convolution-subtraction method (18) with kernels obtained by background tail fitting. This scanner's characteristics and standard performance have been previously reported in detail (5).

Whole-Body PET/CT Scanner

Whole-body PET/CT scans were obtained using a Biograph Horizon TrueV FDG-PET/CT system (Siemens Medical

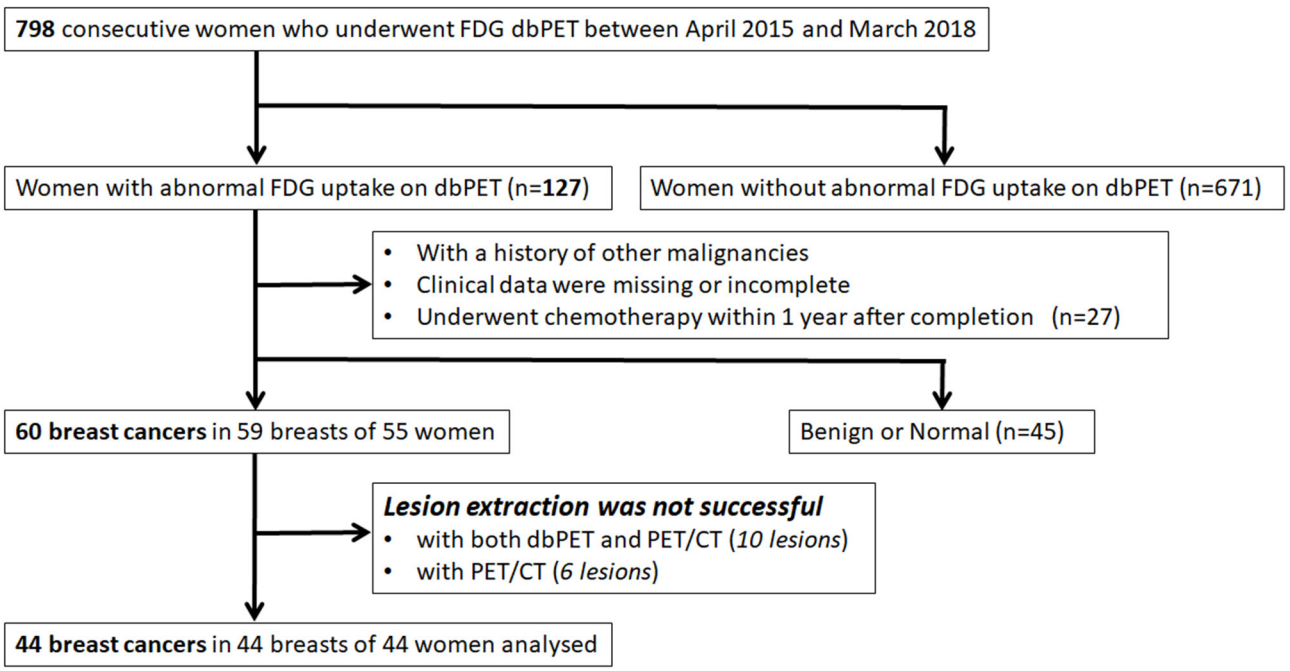


FIGURE 1 | Inclusion and exclusion criteria. FDG, fluorodeoxyglucose; PET, positron emission tomography; dbPET, dedicated breast positron emission tomography; CT, computed tomography.

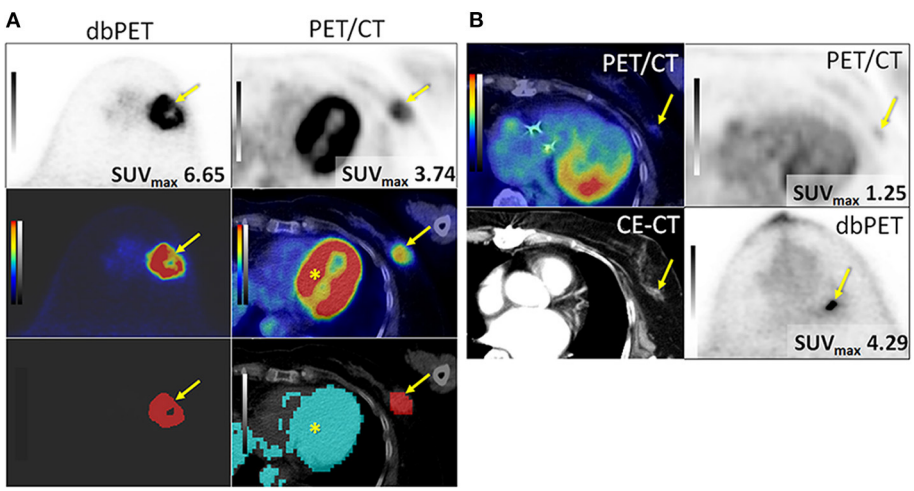


FIGURE 2 | Successfully **(A)** and unsuccessfully **(B)** extracted breast cancers with dbPET and whole-body PET/CT. **(A)** Invasive ductal carcinoma in the left breast of an 84-year-old woman. BC (arrow) was successfully extracted with both dbPET and whole-body PET/CT (two images at the bottom) and clearly separated from the physiological uptake of the myocardium (*) with whole-body PET/CT. **(B)** Invasive ductal carcinoma in the left breast of a 74-year-old woman. It was successfully extracted with dbPET but failed with whole-body PET/CT; therefore, it was excluded from this study. BC, breast cancer; CE-CT, contrast-enhanced computed tomography; dbPET, dedicated breast positron emission tomography.

Solutions, Knoxville, TN, USA). This system has 52 detector rings consisting of 160 blocks. Each block containing an array of 13 × 13 lutetium oxyorthosilicate crystals (4 mm × 4 mm × 20 mm) covering an axial FOV of 221 mm and transaxial FOV of 690 mm. A CT scan was performed for attenuation correction (130 kV; 15–70 mA; tube rotation time, 0.6 s per rotation; pitch, 1; transaxial FOV, 700 mm; and section thickness, 5 mm).

Data Acquisition and Image Reconstruction

All patients fasted for at least 6 h before administration of 18F-FDG (3 MBq/kg). Sixty min after the injection, patients underwent whole-body PET/CT scanning prior to dbPET. The PET/CT images were reconstructed using the ordered subset expectation maximization method and time-of-flight algorithm with 4 iterations and 10 subsets. The CT data were resized from a 512 × 512 matrix to a 180 × 180 matrix to match the PET data and construct CT-based transmission maps for attenuation correction of the PET data with a post-reconstruction smoothing Gaussian filter (5 mm FWHM). The voxel size was 4.11 x 4.11 x 5 mm^3.

Approximately 90 min after FDG injection, after the whole-body PET/CT scan, dbPET scanning was performed for 7 min for each breast. The dbPET images were reconstructed using a three-dimensional list mode dynamic row-action maximum-likelihood algorithm with one iteration and 128 subsets, a relaxation control parameter of $\beta = 20$, and a matrix size in the axial view of 236 × 200 × 236. Reconstruction was done with a post-reconstruction smoothing Gaussian filter (1.17-mm FWHM). Attenuation correction using a uniform attenuation map with object boundaries obtained from the emission data was performed on phantom or clinical dbPET images, respectively. The convolution subtraction method was the scatter correction method used, with kernels obtained by background tailfitting (18). The voxel size was 0.78 × 0.78 × 2.34 mm^3.

Image Analysis

The SUV of each tumor was measured by a spherical volume of interest (VOI). The SUV was a dose- and body-weight-corrected value of tissue tracer concentration. The delineation method used a relative threshold set to 40% of the maximum standardized uptake value (SUVmax) in the lesion to identify the VOI according to a previous report (19). Compared with previous studies, metabolic tumor volume (MTV) and total lesion glycolysis (TLG) were calculated for reference. The MTV was defined as the VOI volume, and TLG was calculated by multiplying the MTV by the mean SUV (SUVmean). The SUVmean was defined as the average of all voxels in the VOI. All processes were performed using Metavol (PMID: 25162396).

Texture Analysis

The SUV was resampled using 64 discrete values from the lowest to highest SUV. We used the PTexture package that we developed in a previous study (20). PTexture is a package using Python to compute texture features from voxel lists. The entire source codes of PTexture are available at https://github.com/metavol/ptexture. Further details regarding texture analysis have been previously reported (21). Texture features were computed only from PET images and not CT because dbPET is not attached to a CT.

Statistical Analysis

Wilcoxon signed-rank test was used for the inter-group comparison of texture features for each BC characteristic. Because of the large number of texture features extracted

TABLE 1 | Patient and tumor characteristics.

	Number of lesions (n = 44) in 44 breasts of 44 patients
Age, years (median, range)	59, 37–87
Tumor (T)-category	
Tis	4
T1	12
T2	25
T3	3
T4	0
Lymph node (N)-category	
N0	22
N1	13
N2	6
N3	3
Stage	
0	4
I	9
II	22
III	8
IV	1
Histology	
Non-invasive ductal	4
Invasive carcinoma	
Ductal	37
Lobular	1
Ductal and lobular	1
Apocrine	1
Ki67 level	
20%>	14
≥ 20%	26
Not specified	4
Tumor subtype	
Luminal A	19
Luminal B/HER2-	15
Luminal B/HER2+	2
HER2	3
Triple-negative	5

HER2, human epidermal growth factor receptor 2 amplified.

from PET components and their high correlation with each other, feature reduction was performed by principal component analysis (PCA). PCA was performed on 38 texture features. Because PCA can extract features without reducing the number of features in advance, all the obtained texture features were used for PCA. The predictive performance of each feature in classifying patients according to Tumor (T)-category, Lymph node (N)-category, molecular subtype, and Ki67 level was evaluated and quantified using the area under the curve (AUC) in receiver-operating characteristic (ROC) analysis. A p-value < 0.05 was considered to indicate statistical significance. JMP®15 (SAS Institute Inc., Cary, NC, USA) was used for the analyses.

RESULTS

Patient Characteristics

Of 798 enrolled consecutive women, 127 had abnormal findings on dbPET, and 60 BCs in 59 breasts of 55 women were histopathologically proven before or within 3 months after PET examination. Of 60 BCs that could be visually detected as showing abnormal FDG uptake with dbPET, 10 could not be identified with whole-body PET/CT. The visual detection rate of BC with PET/CT was 83% (50/60) of dbPET. Of 50 BCs that could be detected by both dbPET and whole-body PET/CT, 6 [ductal carcinoma *in situ* (DCIS): 1; T1b: 4; and T1c: 1] could not be successfully extracted with PET/CT. In addition, 10 (DCIS: 6; T1a: 2; T1b: 1 and T1c: 1) could not be successfully extracted with either dbPET or PET/CT because the lesion-to-background SUV ratio was low. Therefore, they were excluded from this study. Some BCs were false-negative with dbPET even in the FOV, but they could not be detected with whole-body PET/CT either. Finally, 44 BCs in 44 breasts of 44 women with a median age of 59 years (range: 37–87) were included in this study (**Figure 1**). **Figure 2** shows representative BCs that were successfully extracted and failed to be extracted. Forty of these were invasive BCs, and four were non-invasive. Four, 9, 22, 8, and 1 BC patients were diagnosed with BC stage 0, I, II, III, and IV, respectively. Patient and tumor characteristics are shown in **Table 1**.

Comparison of the Ability to Predict Tumor Characteristics Using Texture Features of dbPET and Whole-Body PET/CT

We calculated a total of 38 texture features. Five features were computed from a histogram. Four matrices, comprised of gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level zone size matrix (GLZSM), and neighborhood gray-level difference matrix (NGLDM), were generated. Thirty-one features were computed from these four matrices. The number of voxels and sum of SUV were added to the 36 calculated texture features, and a total of 38 features were finally used in the analysis. These texture features are generally applied to the previous PET studies of various cancers (21). Twenty-three dbPET and 17 PET/CT texture features were significantly different between the Tis-1 and T2-4 groups. In addition, four dbPET and four PET/CT texture features were significantly different between Luminal A and the other groups.

TABLE 2 | Associations (*p-value*) between texture PET parameters and tumor characteristics.

	dbPET		PET/CT	
	Tis-1 vs. T2-4	**Luminal A vs. Others**	**Tis-1 vs. T2-4**	**Luminal A vs. Others**
Num of Voxels	NS	NS	0.0179	NS
SUV_{sum}	NS	NS	0.0047	0.0284
SDhist	<0.0001	NS	NS	NS
Skewness	0.0003	NS	0.0084	NS
Kurtosis	<0.0001	NS	0.0104	0.0152
$Energy_{Hist}$	0.0003	NS	<0.0001	NS
$Entropy_{Hist}$	0.0006	NS	<0.0001	NS
$Homogeneity_{GLCM}$	NS	NS	NS	NS
$Energy_{GLCM}$	NS	NS	0.0157	NS
$Correlation_{GLCM}$	NS	NS	0.0359	NS
$Contrast_{GLCM}$	NS	NS	NS	NS
$Entropy_{GLCM}$	<0.0001	0.0404	0.0157	NS
$Dissimilarity_{GLCM}$	<0.0001	NS	NS	NS
SRE	0.0005	NS	NS	NS
LRE	<0.0001	NS	NS	NS
LGRE	<0.0001	NS	0.009	NS
HGRE	<0.0001	NS	NS	NS
SRLGE	<0.0001	0.0302	0.0073	NS
SRHGE	<0.0001	0.036	NS	NS
LRLGE	NS	NS	0.0097	NS
LRHGE	NS	NS	NS	NS
GLNUr	NS	NS	NS	NS
RLNU	NS	NS	0.0147	NS
RP	NS	NS	NS	NS
SZE	NS	NS	NS	NS
LZE	<0.0001	NS	NS	NS
LGZE	<0.0001	NS	0.0058	NS
HGZE	<0.0001	NS	NS	0.0302
SZLGE	<0.0001	NS	0.0128	NS
SZHGE	<0.0001	0.032	NS	NS
LZLGE	NS	NS	NS	NS
LZHGE	NS	NS	NS	0.0196
GLNUz	NS	NS	NS	NS
ZSNU	0.0073	NS	0.0084	NS
ZP	0.0318	NS	NS	NS
$Coarseness_{NGLDM}$	0.0032	NS	NS	NS
$Contrast_{NGLDM}$	<0.0001	NS	NS	NS
$Busyness_{NGLDM}$	<0.0001	NS	0.0233	NS

Numbers are p-values when there was a significant difference. NS not significant.
SUV, standardized uptake value; SUV_{sum}, sum of SUV; SDhist, standard deviation from a histogram; $Energy_{Hist}$, energy from a histogram; $Entropy_{Hist}$, entropy from a histogram; GLCM, gray-level cooccurence matrix; SRE, short-run emphasis; LRE, long-run emphasis; LGRE, low gray-level run emphasis; HGRE, high gray-level run emphasis; SRLGE, short-run low gray-level emphasis; SRHGE, short-run high gray-level emphasis; LRLGE, long-run low gray-level emphasis; LRHGE, long-run high gray-level emphasis; GLNUr, gray-level non-uniformity for run; RLNU, run-length non-uniformity; RP, run percentage; SZE, short-zone emphasis; LZE, long-zone emphasis; LGZE, low gray-level zone emphasis; HGZE, high gray-level zone emphasis; SZLGE, short-zone low gray-level emphasis; SZHGE, short-zone high gray-level emphasis; LZLGE, long-zone low gray-level emphasis; LZHGE, long-zone high gray-level emphasis; GLNUz, gray-level non-uniformity for zone; ZSNU, zone-size nonuniformity; ZP, zone percentage; NGLDM, neighborhood gray-level different matrix; $Coarseness_{NGLDM}$, coarseness from a NGLDM; $Contrast_{NGLDM}$, contrast from a NGLDM; $Busyness_{NGLDM}$, busyness from a NGLDM.

TABLE 3 | ROC analysis for classification of tumor characteristics by texture features using principal component analysis.

	Tumor characteristic		AUC	Sensitivity	Specificity	Accuracy
dbPET	T-category	T1 vs. T2-4	0.89	0.82	0.94	0.86
	N-category	Negative vs. Positive	0.66	0.85	0.54	0.68
	Molecular subtype	Luminal A vs. Others	0.73	0.52	0.89	0.68
	Ki67 level	20%> vs. 20%≤	0.75	0.72	0.79	0.74
PET/CT	T-category	T1 vs. T2-4	0.94	0.93	0.88	0.91
	N-category	Negative vs. Positive	0.71	0.70	0.75	0.73
	Molecular subtype	Luminal A vs. Others	0.82	0.68	0.84	0.75
	Ki67 level	20%> vs. 20%≤	0.86	0.76	0.86	0.79

ROC, receiver operating characteristic; dbPET, dedicated breast positron emission tomography; PET/CT, positron emission tomography/computed tomography; AUC, area under the curve.

Other texture features of dbPET or PET/CT were not associated with any BC characteristics (**Table 2**).

We decided to use the first PCs with eigenvalues >1. As a result, 5 PCs were used for each of dbPET and whole-body PET/CT, and they explained 94 and 92.2% of the variance in dbPET and whole-body PET/CT, respectively. Scree plots and factor loadings of the PCs in the PCA of texture features are shown in **Supplementary Figure 1** and **Supplementary Table 1**, respectively. The PCs of the 38 textural features obtained from dbPET and PET/CT using PCA were excellent in predicting T-category (AUC = 0.89 and 0.94, respectively) and good to fair in predicting N-category (AUC = 0.66 and 0.71), molecular subtype (AUC = 0.73 and 0.82), and Ki67 levels (AUC = 0.75 and 0.86, **Table 3**). The ROC curves of dbPET and PET/CT, shown in **Figure 3**, were similar, and there was no statistically significant difference in predictive power between them.

Table 4 summarizes the associations between conventional PET parameters and tumor characteristics. MTV and TLG of both dbPET and PET/CT were associated with the T-category. TLG of PET/CT was associated with tumor subtype; however, no other parameters were associated with BC characteristics (**Table 4**).

DISCUSSION

This is the first study comparing the texture features derived from PET images of BC between dbPET and whole-body PET/CT to the best of our knowledge. Although no individual feature among the 38 texture features calculated from dbPET and whole-body PET/CT BC images was significantly associated with all tumor characteristics of interest, the PCs derived by PCA of these texture features obtained with both modalities had good predictive power for T-category, N-category, molecular subtype (Luminal A vs. others), and Ki67 level. Our results suggest that the texture features derived from PET/CT images of histopathologically proven BC, which has enough volume to be successfully extracted, may apply to the evaluation of neoadjuvant chemotherapy and prognosis prediction. However, PET/CT is inferior in spatial resolution to dbPET.

Moscoso et al. have reported that some texture features of dbPET (Dissimilarity, Entropy, Homogeneity, ZP) were associated with tumor size and molecular subtype (22). We calculated and analyzed 38 texture features, including the first- and higher-order statistical features, thus considering spatial position information. Previous experimental and clinical results demonstrated the importance of using higher-order statistical features in texture analysis (23, 24). Our results also demonstrated the good diagnostic ability of texture features for all the tumor characteristics in this study. However, it was impossible to identify specific features as predictors of every tumor characteristic.

Recently, due to the increased interest and number of published studies, several issues on the use of radiomics have emerged. First, because the number of parameters considered has gradually increased, and the analyses have become increasingly complex, it is difficult to determine the most effective features associated with BC characteristics. Second, there is no consensus on the radiomics method using PET images of BC; therefore, the results differ slightly among studies.

In this study, the texture analysis of dbPET could predict BC characteristics with the same accuracy as that of whole-body PET/CT. However, six BCs, all early-stage, were excluded from the whole-body PET/CT analysis as they could not be successfully extracted on PET/CT images. In previous studies comparing dbPET and whole-body PET/CT, the analysis of standard performance differences between the two scanners in phantom tests and some clinical trials have shown the superiority of dbPET over whole-body PET/CT (12, 13). The significance of dbPET may be demonstrated in the diagnosis of early, small BCs. Texture analyses with a large number of early-stage BCs may also show the efficacy of dbPET.

Another issue was that some BCs, even though they could be visually confirmed to be abnormal on dbPET images, were excluded from this study. This issue may suggest a challenge worth future investigation. The issue is to determine the most suitable tumor extraction method (e.g., gradient methods) for dbPET texture analysis instead of the VOI setting's optimal thresholds. The effect of the image reconstruction methods on the texture analysis of dbPET should also be clarified in the future. High-resolution reconstruction of PET/CT images has been reported to change the textural features compared to standard reconstructed clinical PET images (11, 25). It is necessary to assess how the texture features of dbPET change compared to that of the whole-body PET/CT image due to the differences in the reconstruction method.

Our study has several limitations. The primary limitations include the retrospective nature of the study and small cohort. The evaluation of the association between PET parameters and BC characteristics such as molecular subtype, histopathological grade, and Ki-67 expression could, therefore, not be fully

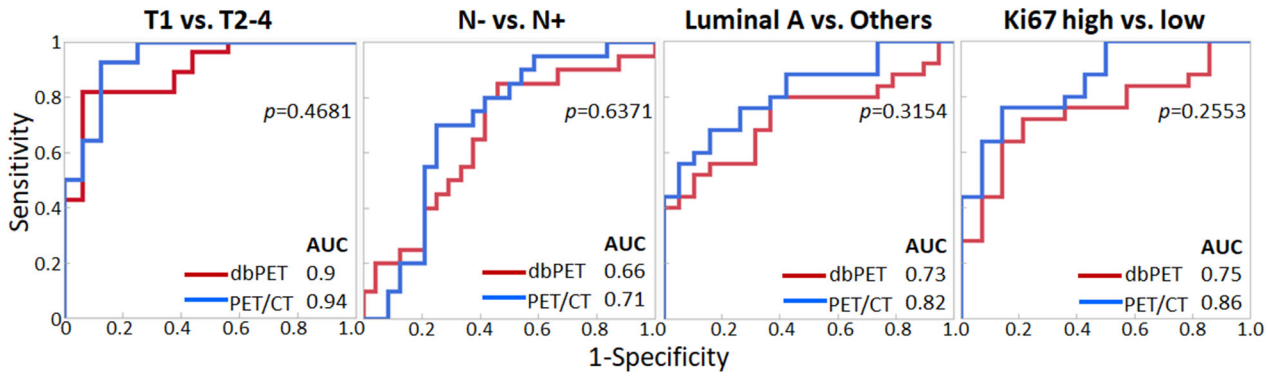


FIGURE 3 | Comparison of ROC analysis for classifying breast cancer characteristics by texture features of dbPET (red line) and whole-body PET/CT (blue line). The AUC values in PCA with texture features of both dbPET and PET/CT were not significantly different. AUC, area under the curve; dbPET, dedicated breast positron emission tomography; CT, computed tomography; PCA, principal component analysis; ROC, receiver-operating-characteristic.

TABLE 4 | Associations (*p-value*) between conventional PET parameters and tumor characteristics.

	dbPET				PET/CT			
Tumor characteristic	**T-category**	**N-category**	**Molecular subtype**	**Ki67 level**	**T-category**	**N-category**	**Molecular subtype**	**Ki67 level**
	Tis-1 vs. 2-4	**Negative vs. Positive**	**Luminal A vs. Others**	**>20% vs. ≤20%**	**Tis-1 vs. T2-4**	**Negative vs. Positive**	**Luminal A vs. Others**	**>20% vs. 20%≤**
SUV_{max}	0.053	0.0886	0.078	0.0901	0.1078	0.3402	0.0666	0.0741
MTV	<0.0001*	0.9295	0.078	0.4309	0.0208*	0.7691	0.2873	0.3797
TLG	<0.0001*	0.456	0.0504	0.2253	0.0077*	0.4484	0.0376*	0.0841

**Statistically significant. dbPET, dedicated breast positron emission tomography; PET/CT, positron emission tomography/computed tomography; SUVmax, maximum standardized uptake value; MTV, metabolic tumor volume; TLG, total lesion glycolysis.*

conducted in the context of an associated prognosis. We tried to apply some machine learning classifiers (Support Vector Machine and Random Forest). However, a classifier with high generalization performance could not be obtained. Further studies following the accumulation of a larger number of clinical cases are needed. With more data, the utility of texture analysis could be further generalized by creating classifiers with higher generalization performance. Second, the images analyzed in this study were acquired at 60 min for whole-body PET/CT and 90 min for dbPET after FDG injection. FDG uptake in BC lesions is known to increase over time, which might affect the results. However, it is difficult to change the order of the scans in a clinical setting; thus, this issue needs future investigations.

In conclusion, dbPET was overall superior for detection of BC. However, for BCs that could be successfully extracted, whole-body PET/CT showed the equivalent predictive ability of tumor characteristics using texture analysis to that of dbPET.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Kofu Neurosurgical Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Material preparation, data collection, and analysis were performed by YS. Image analysis was performed by KH. Statistical analysis was performed by YS, DT, and SF. The first draft of the manuscript was written by YS and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2020.603303/full#supplementary-material

REFERENCES

1. Groheux D, Cochet A, Humbert O, Alberini JL, Hindié E, Mankoff D. 8F-FDG PET/CT for staging and restaging of breast cancer. *J Nucl Med*. (2016) 57:17S−26S. doi: 10.2967/jnumed.115157859
2. Cermik TF, Mavi A, Basu S, Alavi A. Impact of FDG PET on the preoperative staging of newly diagnosed breast cancer. *Eur J Nucl Med Mol Imaging*. (2008) 35:475–83. doi: 10.1007/s00259-007-0580-5
3. Prieto E, Domínguez-Prado I, García-Velloso MJ, Peñuelas I, Richter JÁ, Martí-Climent JM. Impact of time-of-flight and point-spread-function in SUV quantification for oncological PET. *Clin Nucl Med*. (2013) 38:103–9. doi; 10.1097/RLU.0b013e318279b9df
4. Narayanan D, Madsen KS, Kalinyak JE, Berg WA. Interpretation of positron emission mammography: feature analysis and rates of malignancy. *AJR Am J Roentgenol*. (2011) 196:956–70. doi: 10.2214/AJR.104748
5. Miyake KK, Matsumoto K, Inoue M, Nakamoto Y, Kanao S, Oishi T, et al. Performance evaluation of a new dedicated breast PET scanner using NEMA NU4-2008 standards. *J Nucl Med*. (2014) 55:1198–203. doi: 10.2967/jnumed.113131565
6. Satoh Y, Motosugi U, Omiya Y, Onishi H. Unexpected abnormal uptake in the breasts at dedicated breast PET: incidentally detected small cancers or nonmalignant features? *AJR Am J Roentgenol*. (2019) 212:443–9. doi: 10.2214/AJR.1820066
7. Chitalia RD, Kontos D. Role of texture analysis in breast MRI as a cancer biomarker: a review. *J Magn Reson Imaging*. (2019) 49:927–38. doi: 10.1002/jmri26556
8. Klimonda Z, Karwat P, Dobruch-Sobczak K, Piotrzkowska-Wróblewska H, Litniewski J. Breast-lesions characterization using quantitative ultrasound features of peritumoral tissue. *Sci Rep*. (2019) 9:7963. doi: 10.1038/s41598-019-44376-z
9. Orlhac F, Soussan M, Maisonobe JA, Garcia CA, Vanderlinden B, Buvat I. 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*. (2014) 55:414–22. doi: 10.2967/jnumed.113129858
10. Ha S, Park S, Bang JI, Kim EK, Lee HY. Metabolic radiomics for pretreatment 18F-FDG PET/CT to characterize locally advanced breast cancer: histopathologic characteristics, response to neoadjuvant chemotherapy, and prognosis. *Sci Rep*. (2017) 7:1556. doi: 10.1038/s41598-017-01524-7
11. Aide N, SalomonT, Blanc-Fournier C, Grellard JM, Levy C, Lasnon C. Implications of reconstruction protocol for histo-biological characterisation of breast cancers using FDG-PET radiomics. *EJNMMI Res*. (2018) 8:114. doi: 10.1186/s13550-018-0466-5
12. Nishimatsu K, Nakamoto Y, Miyake KK, Ishimori T, Kanao S, Toi M, et al. Higher breast cancer conspicuity on dbPET compared to WB-PET/CT. *Eur J Radiol*. (2017) 90:138–45. doi: 10.1016/j.ejrad.2017.02046
13. Satoh Y, Motosugi U, Imai M, Onishi H. Comparison of dedicated breast positron emission tomography and whole-body positron emission tomography/computed tomography images: a common phantom study. *Ann Nucl Med*. (2020) 34:119–27. doi: 10.1007/s12149-019-01422-0
14. Amin MB, Edge S, Greene F, Byrd D, Brookland RK, Washington MK, editors, et al. *AJCC Cancer Staging Manual*. New York: Springer (2018).
15. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol*. (2013) 24:2206–23. doi: 10.1093/annonc/mdt303
16. Qi J, Kuo C, Huesman RH, Klein GJ, Moses WW, Reutter BW. Comparison of rectangular and dual-planar positron emission mammography scanners. *IEEE Trans Nucl Sci*. (2002) 4:2089–96. doi: 10.1109/TNS.2002803810
17. Yamakawa Y, Kitamura K. Attenuation correction using level set method for application specific PET scanners. *IEEE Nucl Sci Symp Conf Rec*. (2011) 3130–2. doi: 10.1109/NSSMIC.20116152569
18. Bailey DL, Meikle SR. A convolution-subtraction scatter correction method for 3D PET. *Phys Med Biol*. (1994) 39:411–24. doi: 10.1088/0031-9155/39/3/009
19. Son SH, Kim D.-H, Hong CM, Kim C.-Y, Jeong SY, Lee SW, et al. Prognostic implication of intratumoral metabolic heterogeneity in invasive ductal carcinoma of the breast. *BMC Cancer*. (2014) 14:585. doi: 10.1186/1471-2407-14-585
20. Manabe O, Ohira H, Hirata K, Hayashi S, Naya M, Tsujino I, et al. Use of 18F-FDG PET/CT texture analysis to diagnose cardiac sarcoidosis. *Eur J Nucl Med Mol Imaging*. (2019) 46:1240–7. doi: 10.1007/s00259-018-4195-9
21. Ha S, Choi H, Paeng JC, Cheon GJ. Radiomics in oncological PET/CT: methodological overview. *Nucl Med Mol Imaging*. (2019) 53:14–29. doi: 10.1007/s13139-019-00571-4
22. Moscoso A, Ruibal Á, Domínguez-Prado I, Fernández-Ferreiro A, Herranz M, Albaina L, et al. Texture analysis of high-resolution dedicated breast 18 F-FDG PET images correlates with immunohistochemical factors and subtype of breast cancer. *Eur J Nucl Med Mol Imaging*. (2018) 45:196–206. doi: 10.1007/s00259-017-3830-1
23. Harrabi R, Sayadi M, Fnaiech F. Higher order statistics applied to image segmentation. *IEEE Nucl Sci Symp Conf Rec*. (2009). doi: 10.1109/IECON.2009.5415191
24. Soussan M, Orlhac F, Boubaya M, Zelek L, Ziol M, Eder V, et al. Relationship between tumor heterogeneity measured on FDG-PET/CT and pathological prognostic factors in invasive breast cancer. *PLoS ONE*. (2014) 9:e94017. doi: 10.1371/journal.pone0094017
25. Shiri I, Rahmim A, Ghaffarian P, Geramifar P, Abdollahi H, Bitarafan-Rajabi A. The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies. *Eur Radiol*. (2017) 27:4498–509. doi: 10.1007/s00330-017-4859-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 13 January 2021
doi: 10.3389/fmed.2020.617173

Exploring the Pattern Associated With Longitudinal Changes of β-Amyloid Deposition During Cognitively Normal Healthy Aging

Yunyan Xie[1†], Qin Yang[1†], Chunhua Liu[2], Qi Zhang[2], Jiehui Jiang[2,3,4], Ying Han[1,5,6*] and the Alzheimer's Disease Neuroimaging Initiative*

[1] Department of Neurology, Xuanwu Hospital of Capital Medical University, Beijing, China, [2] Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, School of Communication and Information Technology, Shanghai University, Shanghai, China, [3] Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai, China, [4] Institute of Biomedical Engineering, Shanghai University, Shanghai, China, [5] Center of Alzheimer's Disease, Beijing Institute for Brain Disorders, Beijing, China, [6] National Clinical Research Center for Geriatric Disorders, Beijing, China

OPEN ACCESS

Edited by:
Xiu Ying Wang,
The University of Sydney, Australia

Reviewed by:
Susanne Asenbaum-Nan,
Lower Landeskliniken-Holding, Austria
Zhen Cheng,
Stanford University, United States

***Correspondence:**
Jiehui Jiang
jiangjiehui@shu.edu.cn
Ying Han
hanying@xwh.ccmu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: 14 October 2020
Accepted: 14 December 2020
Published: 13 January 2021

Citation:
Xie Y, Yang Q, Liu C, Zhang Q, Jiang J, Han Y and the Alzheimer's Disease Neuroimaging Initiative (2021) Exploring the Pattern Associated With Longitudinal Changes of β-Amyloid Deposition During Cognitively Normal Healthy Aging. Front. Med. 7:617173. doi: 10.3389/fmed.2020.617173

The aim of this study was to determine a pattern associated with longitudinal changes of β-amyloid (Aβ) deposition during cognitively normal(CN) healthy aging. We used ^{18}F-florbetapir (AV-45) PET images of the brains of 207 cognitively normal subjects (CN1), obtained through the Alzheimer's Disease Neuroimaging Initiative (ADNI), to identify the healthy aging pattern and 76 cognitively normal healthy subjects (CN2), obtained through the Xuanwu Hospital of Capital Medical University, Beijing, China, to verify it. A voxel-based correlation analysis of standardized uptake value ratio (SUVR) map image and age was conducted using the DPABI (Data Processing & Analysis of Brain Imaging) software to identify the pattern. The sum of squares due to errors (SSE), R-square (R^2) and the root-mean-square error (RMSE) were calculated to assess the quality of curve fitting. Among them, R^2 was proposed as the coherence coefficient, which was as an index to assess the correlation between SUVR value of the pattern and subjects' age. The pattern characterized by age-associated longitudinal changes of Aβ deposition was mainly distributed in the right middle and inferior temporal gyrus, the right temporal pole: middle temporal gyrus, the right inferior occipital gyrus, the right inferior frontal gyrus (triangular portion), and the right precentral gyrus. There were a significant positive correlation between the SUVR value of the pattern and age for each CN group (CN1: $R^2 = 0.120$, $p < 0.001$ for quadratic model; CN2: $R^2 = 0.152$, $p = 0.002$ for quadratic model). These findings suggest a pattern of changes in Aβ deposition that can be used to distinguish physiological changes from pathophysiological changes, constituting a new method for elucidating the neuropathological mechanism of Alzheimer's disease.

Keywords: healthy aging, ^{18}F-AV-45 PET, β-amyloid deposition, brain, pattern

INTRODUCTION

Brain aging, which is influenced by various pathological and psychosocial factors (1), comprises two categories: healthy and pathological aging. According to clinical neurology, healthy aging is defined as "*the cognitively normal (CN) subjects who maintain their normal cognitive level and ability of daily living as they grow older, without neurological diseases*" (2). Pathological aging, which is characterized by the accumulation of extracellular Aβ deposition (3), is considered a major pathological element of Alzheimer's disease (AD) (4). However, the presence of Aβ in the AD brain may also signal a physiological age-associated phenomenon depending on its extent and distribution pattern (5–8). Research evidence suggests that Aβ deposition occurs in the brains of cognitively normal older individuals (9–12). The prevalence of the amyloid burden among cognitively normal older individuals has been estimated to be more than 25% than younger individuals based on the findings of autopsy studies (9, 13, 14). However, given limited knowledge regarding the extent and distribution of Aβ deposition during the healthy aging process, an assessment of changes in Aβ deposition with age is essential for advancing understanding of healthy aging.

Some studies that have measured amyloid deposition in the course of normal aging found a significant linear increase in global Aβ deposition with age (3, 12, 15). One study found a highly significant correlation between increasing age and a reduction in Aβ turnover rates (16). Significant linear increases with age have been observed in the precuneus, temporal cortex, and the anterior and posterior cingulate (3) as well as in the frontal, cingulate and parietal areas, with primary sensory/visual areas being relatively protected from amyloid deposition (17). The findings of the above studies indicate that there may be a linear pattern of brain aging associated with changes in Aβ deposition during healthy aging in cognitively normal adults. However, all above studies were based on western datasets and the repeatability of results was not verified among different ethnic cohorts.

^{18}F-florbetapir (AV-45) is a safe tracer demonstrating high levels of sensitivity and specificity for Aβ detection (18). Aβ deposition in the brain can be quantified within a clinical environment through positron emission computed tomography (PET) scans conducted with ^{18}F-AV-45 (19). Moreover, this technique can be used to study Aβ distribution *in vivo*, enabling the formation and progression of Aβ aggregates in the brain to be monitored (20, 21). Thus, there were two main objectives in this study: (1) to explore a pattern associated with longitudinal changes of Aβ deposition during healthy aging using ^{18}F-AV-45 PET images to quantify Aβ deposition *in vivo*. (2) to verify the repeatability of the healthy aging pattern among western and Chinese cohorts.

MATERIALS AND METHODS

Materials

Two cohorts of neuroimaging data were collected from two independent centers: Cohort A (N = 207, right-handed, CN1) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/) and cohort B (N = 76, right-handed, CN2) from the Department of Neurology, Xuanwu Hospital of Capital Medical University. Both clinical (Sex, Age, Education, Mini-Mental State Examination (MMSE)) and image (18F-AV-45 PET and MRI image) information were selected for the two cohorts. Montreal Cognitive Assessment (MoCA) and Clinical Dementia Rating Sum of Boxes (CDR-SB) were also selected for Cohort A. Notablly, in total 378 images were included for Cohort A because part of subjects had more than one scan (1.83 ± 0.83 scan times per subject), while each subject had only one scan in Cohort B.

Figure 1 shows the exclusion and inclusion criteria applied to CN subjects of Cohort A. The following inclusion criteria were applied: (1) subjects had no history of stroke, hypertension, brain disease, or mental illness. (2) The PET scan results of individuals were Aβ-negative (Aβ-), with a cerebral-to-whole cerebellar florbetapir SUVR value below 1.18 (22). (3) Mini-Mental State Examination (MMSE) scores for individuals were above or equal to 28, and their Clinical Dementia Rating Sum of Boxes (CDR-SB) scores were all 0. Inclusion criteria for subjects in Cohort B were consistent with those for Cohort A. This study was approved by the institutional review boards of ADNI and the Research Ethics Committee of Xuanwu Hospital, Beijing, China.Written, informed consent had been obtained from each subject.

Image Acquisition Protocol

The process of acquiring data for the CN1 group is described in detail in the imaging protocol column of the ADNI database (http://adni.loni.usc.edu/). PET and T1 MRI data were simultaneously obtained for each participant in the CN2 group. All of the participants were invited to undergo optional ^{18}F-florbetapir (AV-45) PET scans in the three-dimensional acquisition mode. A dynamic scan, lasting 35 min, was performed approximately 40 min after participants received an intravenous injection of 7–10 mCi [^{18}F] florbetapir. The PET scan images were analytically reconstructed using a time-of-flight ordered subset expectation maximization (TOF OSEM) algorithm with the following parameters: eight iterations, 32 subsets matrix = 192 × 192, field of view (FOV) = 350 × 350, half-width height = 3.

Three-dimensional T1-weighted magnetization-prepared rapid gradient echo scans were performed using an integrated TOF-capable PET/MR 3.0T imaging device (SIGNA PET/MR, GE Healthcare, Milwaukee, Wisconsin, USA) available at the Xuanwu Hospital of Capital Medical University. The following parameters were applied: SPGR sequence, FOV = 256 × 256 mm^2, matrix = 256 × 256, slice thickness = 1 mm, gap = 0, slice number = 192, repetition time (TR) = 6.9 ms, echo time (TE) = 2.98 ms, inversion time (TI) = 450 ms, flip angle = 12°, voxel size = 1 × 1 × 1 mm^3.

Image Preprocessing

All ^{18}F-AV-45 PET scan images and corresponding T1 images were preprocessed using statistical parametric mapping software (SPM12; https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) in MATLAB (Version R2014a; MathWorks, Natick, MA, United States). We first used the realigning method to ensure

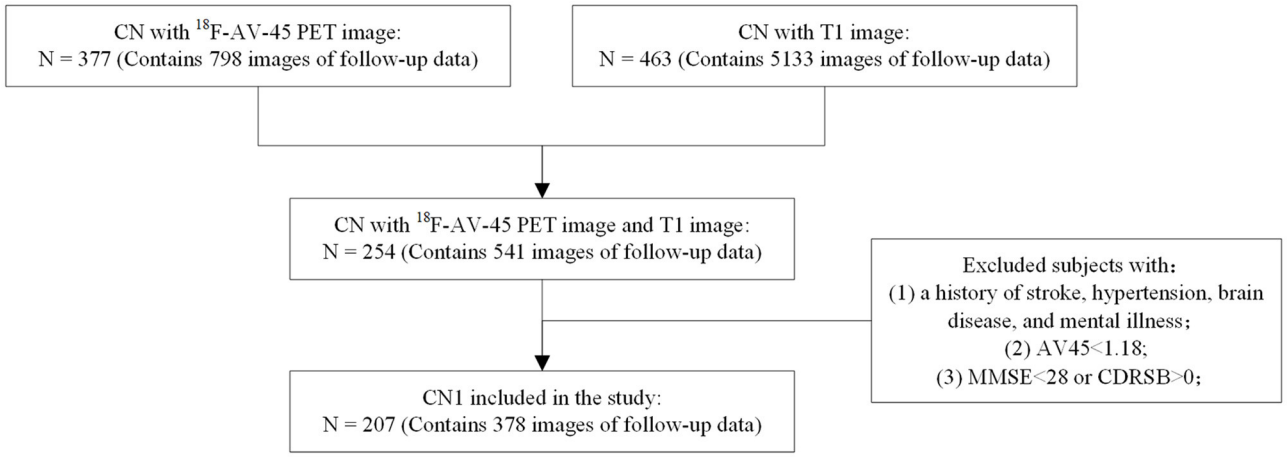


FIGURE 1 | Inclusion and exclusion criteria applied to the ADNI data.

that all frames in the dynamic scans were motion-corrected to the first frame and processed the output single average functional image, reducing system, or head motion errors. Next, we performed a voxel-based partial volume effect (PVE) correction of the functional image using the Müller-Gärtner method (MG), with parameters of white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) obtained through T1 image segmentation. Then, PVE-corrected image was then normalized with reference to the standard Montreal Neurological Institute (MNI) brain space using the deformation field from the MRI image to the MNI space and smoothed to reduce noise and improve image quality using an isotropic Gaussian smoothing kernel with a gaussian filter of 8 mm full-width at half-maximum (FWHM). Lastly, the smoothed functional image was intensity normalized to the mean uptake of whole cerebellum to obtain SUVR map image.

Voxel-Wise PET Analysis

To explore the effect of age on Aβ deposition in the brains of cognitively normal subjects, a voxel-wise correlation analysis of SUVR map iamges was conducted for CN1, with age applied as the seed series and GM, sex, and years of education considered as the covariates. The DPABI software in MATLAB R2014a was used for the analysis. Accordingly, we obtained a statistical map (false discovery rate (FDR) corrected with $q < 0.01$) reflecting the change trend and degree of Aβ deposition in the aging brain. Thereby voxels relating to aging were obtaining with the absolute value of the correlation coefficient ≥ 0.3. As a final step, we mapped the voxels on to the MNI standard space to obtain statistical brain regions as ROIs. To verify that ROIs actually reflect the effect of aging on Aβ deposition in cognitively normal individuals, we examined the correlations between the SUVR values of the healthy aging pattern and age for individuals in the CN1 group and compared the results with the SUVR value for the whole brain. SUVR values were plotted against subjects' ages and fitted using three separate models, namely a linear model:

$$y = at + b, \tag{1}$$

a quadratic model:

$$y = at^2 + bt + c, \tag{2}$$

and an exponential model:

$$y = ae^{bt} + c, \tag{3}$$

where t denotes age and a, b, and c are the parameters to be estimated from the data plotted for the SUVR values and ages of subjects in the CN1 group. The sum of squares due to errors (SSE), R-square (R^2), and the root-mean-square error (RMSE) values were calculated to assess the quality of fit. Among them, R^2 was proposed as the coherence coefficient, which was as an index to assess the correlation between SUVR value of the pattern and subjects' age. Subsequently, the model with the best quality of fit was assigned to the plotted SUVR and age obtained for the CN1 group to evaluate the change trend of Aβ deposition with aging. Forward validation was performed on the CN2 group.

Statistical Analysis

The quantitative results obtained with MATLAB were subjected to a statistical analysis using the SPSS software, version 18.0 (SPSS Inc., IBM Corporation, Chicago, USA). A two-sample t-test was performed to examine differences in continuous variables, and a Chi-square test was conducted to assess categorical variables. $p < 0.05$ was considered statistically significant. The Gramm toolbox in MATLAB was used for plotting and visualizing all of the statistical data presented in this paper (23).

TABLE 1 | Demographic and clinical characteristics of participants.

	CN1	CN2	*P*-value
N	378	76	–
Age (years)	74.8 ± 5.6	65.2 ± 5.2	<0.001
Sex (F/M)	192/186	48/28	0.049
Education (years)	16.6 ± 2.5	12.8 ± 3.4	<0.001
MMSE	29.4 ± 0.8	29.2 ± 0.7	0.119
MoCA	26.2 ± 2.3	–	–
CDR-SB	0.0 ± 0.0	–	–

MMSE, mini-mental state examination; MoCA, montreal cognitive assessment; CDR-SB, clinical dementia rating sum of boxes.

RESULTS

Demographic Characteristics of the Participants

Cohort A contained 378 time ponts of 207 CN subjects (1.83 ± 0.83 time points per subject). Cohort B included 76 time poits of 76 CN subjects (one time point per subject). **Table 1** shows the demographic and clinical details of the two cohorts. As shown in **Table 1**, significant differences between cohort A and cohort B are observed in age ($p < 0.001$), sex ($p = 0.0488$) and education ($p < 0.001$). A slight difference is observed in sex($p = 0.049$) and no significant difference in MMSE ($p = 0.119$). Considering the impact of brain atrophy and differences relating to the sex and education levels of the participants, we reported the results obtained after regressing the covariates of GM, sex, and the number of years of education.

Voxel-Wise PET Analysis

Healthy Aging Pattern

The health aging pattern was identified in CN1 group. The results of the correlation analysis revealed that there was a healthy aging pattern characterized by age-associated longitudinal changes of Aβ deposition was mainly distributed in the right middle and inferior temporal gyrus, the right temporal pole: middle temporal gyrus, the right inferior occipital gyrus, the right inferior frontal gyrus (triangular portion), and the right precentral gyrus (**Figure 2**). No areas of the brain evidenced significantly decreased Aβ deposition (see **Table 2** for details).

Pattern Validation

The health aging pattern was further validated in CN2 group. Following the regression of the covariates of GM, sex, and years of education, SUVR value of the pattern showed a significant positive correlation with age (**Figure 3**), whereas SUVR value of global brain showed a weaker positive correlation with age (**Figure 3**) in the CN1 group. **Table 3** shows the curve fit results for the SUVR of the pattern and age of the three models. Specific results were as follows: SSE = 18.549, $R^2 = 0.118$, and RMSE = 0.222 for the linear model; SSE = 18.505, $R^2 = 0.120$, and RMSE = 0.222 for the quadratic model; and SSE = 18.592, $R^2 = 0.116$, and RMSE = 0.222 for the exponential model. The curve fit results for the SUVR of global brain and age for the three models were as follows: SSE = 12.459, $R^2 = 0.018$, and RMSE = 0.182 for the linear model; SSE = 12.452, $R^2 = 0.019$, and RMSE = 0.182 for the quadratic model; and SSE = 12.461, $R^2 = 0.018$ and RMSE = 0.182 for the exponential model.

Following the regression of the covariates of GM, sex, and educational years, the SUVR value of the pattern showed a significant positive correlation with age (**Figure 4**), whereas SUVR value of global brain showed no significant correlation with age (**Figure 4**) in the CN2 group. **Table 4** shows the curve fit results for SUVR of the pattern and age for the three models. Specific results were as follows: SSE = 0.534, $R^2 = 0.127$ and RMSE = 0.085 for the linear model; SSE = 0.526, $R^2 = 0.152$ and RMSE = 0.085 for the quadratic model; SSE = 0.535, $R^2 = 0.136$, and RMSE = 0.085 for the exponential model. The results for the curve fit of the SUVR of global brain and age were as follows: SSE = 0.502, $R^2 = 0.010$, and RMSE = 0.082 for the linear model; SSE = 0.501, $R^2 = 0.011$, and RMSE = 0.083 for the quadratic model; and SSE = 0.502, $R^2 = 0.010$, and RMSE = 0.082 for the exponential model.

DISCUSSION

During the healthy aging of cognitively normal adults, SUVR value of a healthy aging pattern increased significantly. The pattern was mainly distributed in the right middle and inferior temporal gyrus, the right temporal pole: middle temporal gyrus, the right inferior occipital gyrus, the right inferior frontal gyrus (triangular portion), and the right precentral gyrus. A weak positive correlation was found between SUVR of global brain and age for the CN1 group, with no significant correlation existing for the CN2 group. These results indicate that during the healthy aging process of cognitively normal people, the increase in Aβ deposition is concentrated in specific brain regions rather than being distributed throughout the brain. In addition, the pattern shows a characteristic of asymmetric amyloid accumulation. Alteration in hemispheric asymmetry has been referenced in studies of healthy aging (24, 25). The Right hemi-aging model proposes that the right hemisphere presents greater aging than the left hemisphere (26, 27), so our results were consistent with previous studies.

The best fit curves for the SUVR value of the healthy aging pattern and age within each CN group reflected a change trend of increasing Aβ deposition on the pattern with increasing age and a subsequent decrease in the growth rate of Aβ deposition.

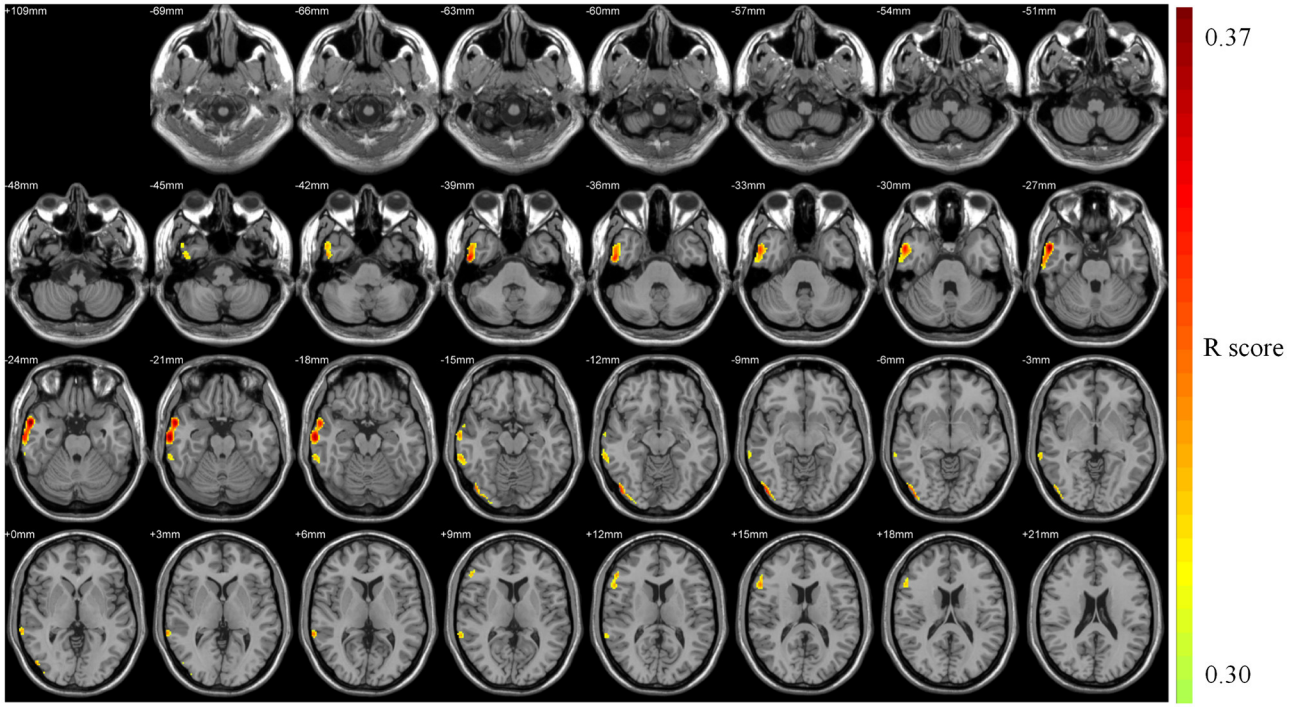


FIGURE 2 | Pattern associated with longitudinal changes of Aβ deposition in cognitively normal adults (CN1) during the healthy aging. The red areas are those with significant increases in Aβ deposition associated with advancing age. No voxels were found that indicated significant decreases in Aβ deposition associated with advancing age.

TABLE 2 | Pattern associated with longitudinal changes of Aβ deposition in cognitively normal adults (CN1) during healthy aging.

Region	Laterality	Cluster extent	Peak Voxel			
			T	X	Y	Z
Middle temporal gyrus	Right	870	0.37	58	4	−24
Inferior temporal gyrus	Right					
Temporal pole: middle temporal gyrus	Right					
Middle temporal gyrus	Right	229	0.33	68	−42	6
Inferior temporal gyrus	Right					
Inferior occipital gyrus	Right	159	0.34	48	−82	−10
Inferior frontal gyrus (triangular portion)	Right	117	0.34	56	20	14
Precentral gyrus	Right	133	0.34	54	10	30

This indicated significant changes in Aβ deposition in aging adults with normal cognition. However, when accumulated deposits of Aβ exceeded a certain threshold, leading to cognitive impairment, aging had weaker effect on Aβ deposition. This finding suggests that the baseline level of Aβ may differ for patients with AD. Moreover, as indicated by the findings of other studies, with the advancement of pathological conditions, Aβ deposition may reach a saturation point and will no longer exhibit a linear relationship with age (17, 28). This finding is supported by that of another study, which revealed that Aβ increases significantly in individuals with normal cognitive functions but that the rate of increase of Aβ slows down following the onset of cognitive impairment (16).

Considering our results together with the findings reported in the literature, we posit that the healthy aging pattern associated with longitudinal changes of Aβ deposition and characteristic regions associated with AD partially overlap, mainly including the middle and inferior temporal gyrus (29, 30) (**Figure 5**). The temporal lobe and occipital cortex are associated with auditory and visual functions (31). The rapid deposition of Aβ in the middle and inferior occipital gyrus and in the middle and inferior temporal gyrus may be one of the reasons why the

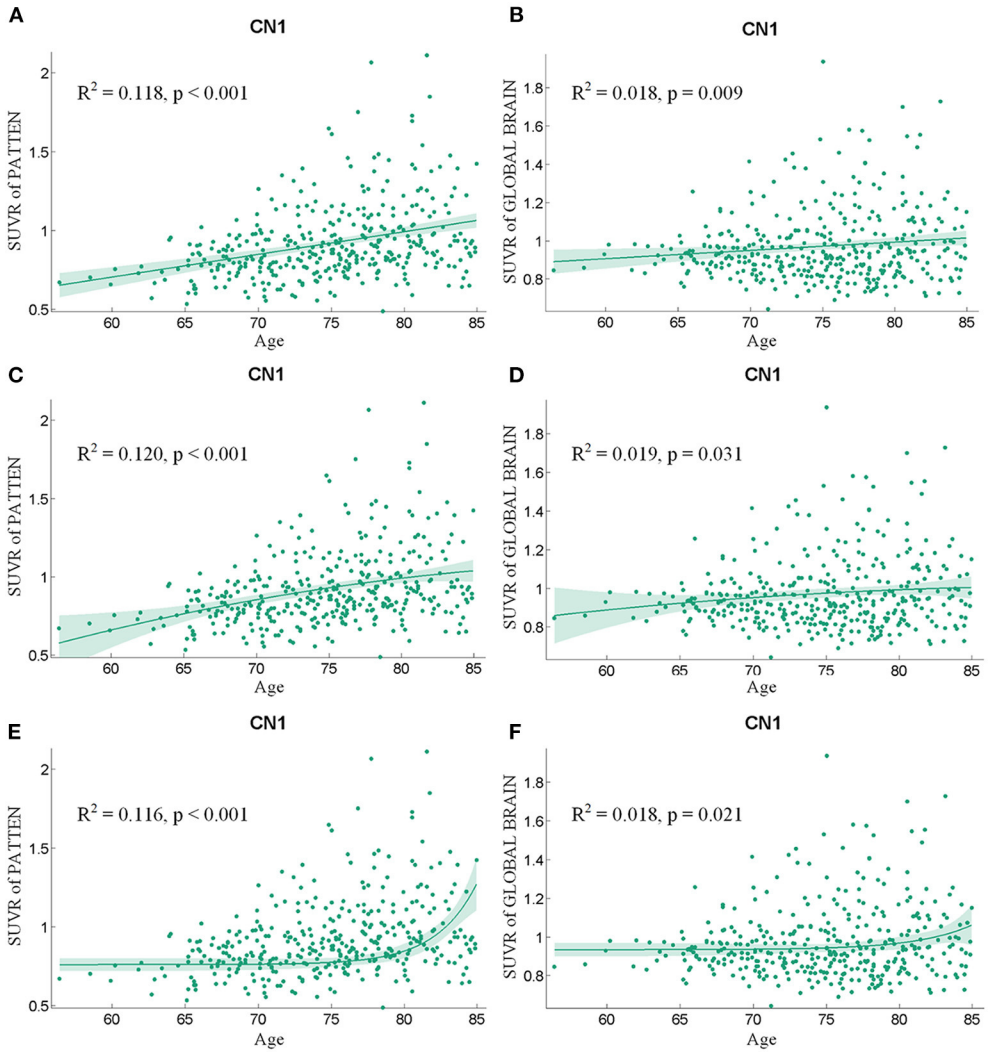


FIGURE 3 | The findings of a correlation analysis of age and the SUVR values of the pattern **(A,C,E)** and the whole brain **(B,D,F)** of individuals in the CN1 group using linear, quadratic, and exponential model fitting, respectively.

auditory and visual fields are influenced by age-associated and neurochemical factors (32) and may reflect a decline in the multisensory integration capacity of older individuals (33). It may account for the importance of age as an influencing factor affecting the diagnosis of AD or early AD.

Moreover, our results on age-associated changes of Aβ deposition could be explained from the perspective of molecular cytology. During the aging process, over-activated microglia may release neurotoxic molecules and pro-inflammatory cytokines, leading to neuronal death and inflammation and an accelerated process of Aβ deposition and accumulation (34, 35). Age-associated increases in microglial activation may contribute to the age-associated increase on Aβ deposition. Past research showed that significant age-associated increases in the total numbers of activated IL-1α+ microglia occurred in mesial temporal lobe (36). And the density of amyloid plaques in the temporal lobe is not related to memory level (37). Our findings on the healthy aging pattern were consistent with previous studies, suggesting that the increase in Aβ deposition promoted by normal aging in the temporal lobe were not caused from cognitive decline.

It should be noted that this study had some limitations. First, the datasets used for the study were limited. Although the pattern were identified and validated using data sourced from the ADNI and Xuanwu Hospital, multicenter research and autopsy results are required to confirm their universality. Second, age differences could be observed between the ADNI and Xuanwu Hospital in this study, and the average age of subjects from Xuanwu hospital was 9.4 years younger than ADNI. Although the correlation between SUVR of pattern and age were found in both cohorts, whether this correlation exsited in older Chinese CN population need be verified in the future. Third, this study was evidently a cross-sectional study, although follow-up data was available for its subjects. A longitudinal study should also be conducted in the

TABLE 3 | The curve-fitting characteristics of SUVR value and age for the CN1 group.

	SUVR of the pattern			SUVR of global brain		
	Linear model	**Quadratic model**	**Exponential model**	**Linear model**	**Quadratic model**	**Exponential model**
SSE	18.549	18.505	18.592	12.459	12.452	12.461
R^2	0.118	0.120	0.116	0.018	0.019	0.018
RMSE	0.222	0.222	0.222	0.182	0.182	0.182
p-value	<0.001	<0.001	< 0.001	0.009	0.031	0.021

SSE, sum of squares due to errors; R^2 R-square; RMSE, the root-mean-square error.

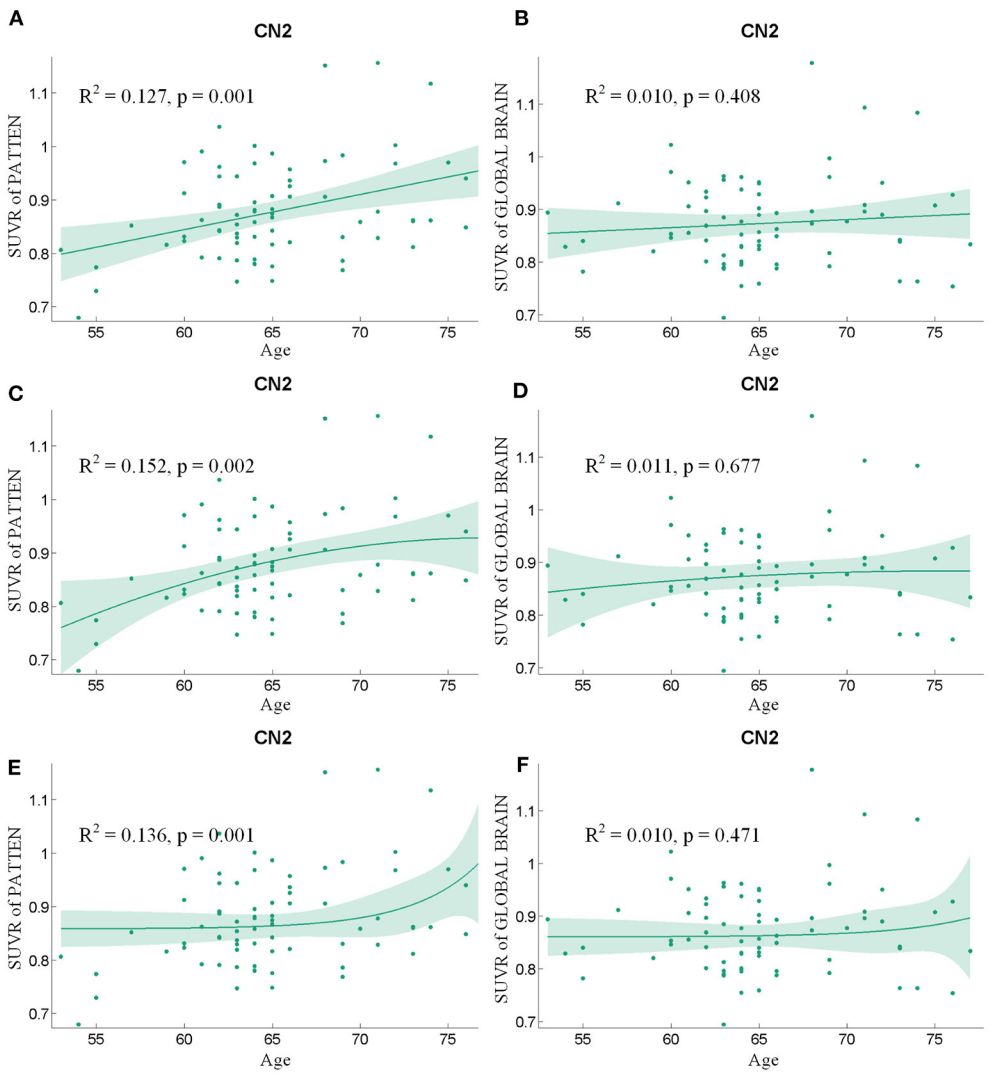


FIGURE 4 | The findings of a correlation analysis of age and the SUVR values of the pattern **(A,C,E)** and the whole brain **(B,D,F)** of individuals in the CN2 group using linear, quadratic, and exponential model fitting respectively.

future. In addition, the similarities and differences between AD-associated and healthy aging patterns merit further study. Finally, although we chose the entire cerebellum as the reference region for calculating the SUVR values, the selection of the reference region has long been a methodologically challenging issue within studies entailing PET imaging analysis. Future comparative studies of different reference regions are urgently required to develop a comprehensive understanding of this compound.

TABLE 4 | The curve-fitting characteristics of SUVR value and age for the CN2 group.

	SUVR of the pattern			SUVR of global brain		
	Linear model	**Quadratic model**	**Exponential model**	**Linear model**	**Quadratic model**	**Exponential model**
SSE	0.534	0.526	0.535	0.502	0.501	0.502
R^2	0.127	0.152	0.136	0.010	0.011	0.010
RMSE	0.085	0.085	0.085	0.082	0.083	0.082
p-value	0.001	0.002	0.001	0.408	0.677	0.471

SSE, sum of squares due to errors; R^2 R-square; RMSE, the root-mean-square error.

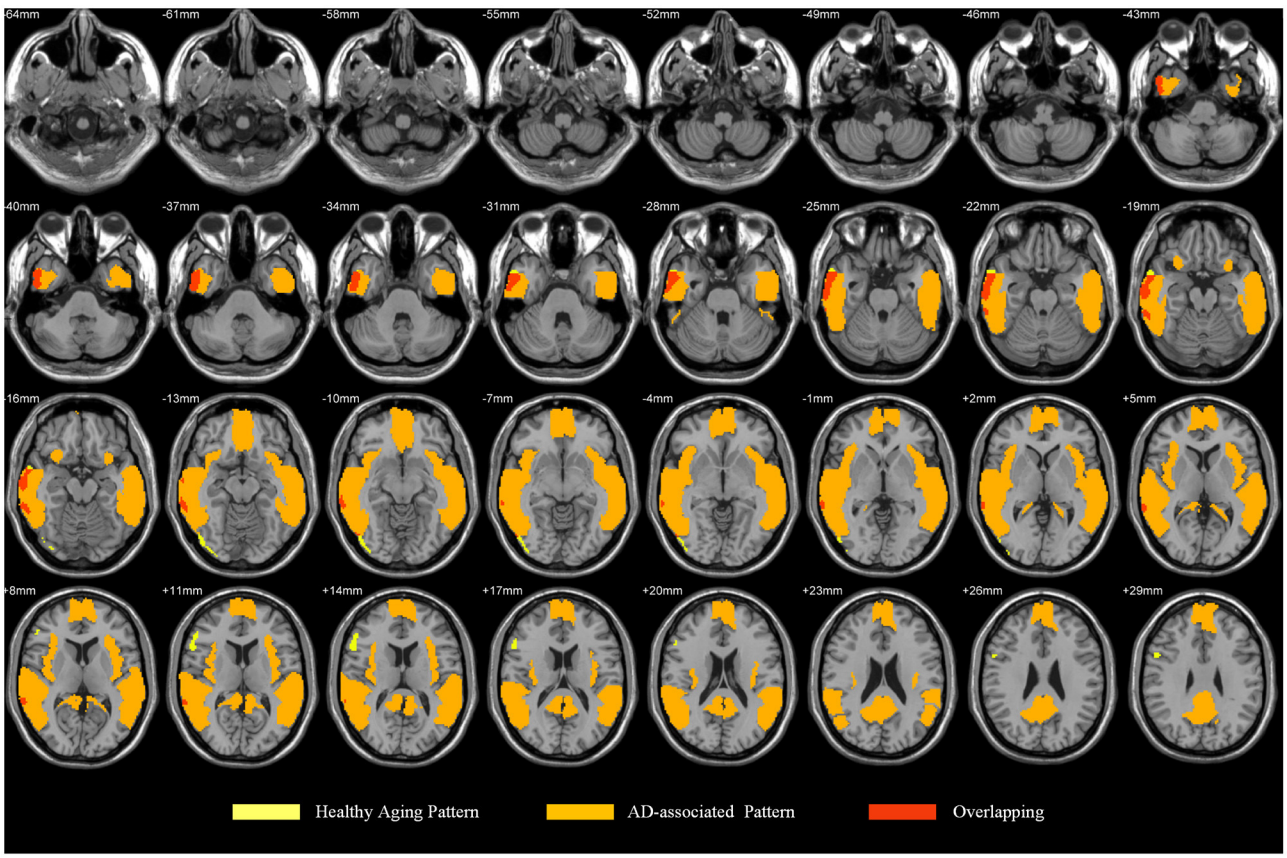


FIGURE 5 | The overlap between our proposed pattern and AD parttern in previous studies.

CONCLUSION

In summary, we performed a voxel-wise correlation analysis to identify a pattern associated with changes in β-amyloid deposition in cognitively normal adults during healthy aging. An assessment of the pattern advances understanding of processual changes entailed in brain aging. The changes in Aβ deposition associated with healthy aging that are reflected in age-associated longitudinal changes of Aβ deposition on specific brain regions are indicative of opportunities for diagnosis and strategies for decelerating aging. More generally, this study may reveal a pattern of changes in Aβ deposition that can be used to distinguish physiological changes from pathophysiological ones.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional review boards of ADNI and the Research Ethics Committee of Xuanwu Hospital, Beijing,

China. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YX and QY are responsible for collecting data, designing experiment and writing article. CL and QZ are responsible for the experimental coding. JJ and YH are responsible for the guidance of the experiment and paper. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by grants received from the National Natural Science Foundation of China (grant numbers 61633018, 82020108013, 61603236, and 81830059); the National Key Research and Development Program of China (grant numbers 2016YFC1306300, 2018YFC1312000, and 2018YFC1707704); the 111 Project (grant number D20031); the Shanghai Municipal Science and Technology Major Project (grant number 2017SHZDZX01); and the Beijing Municipal Commission of Health and Family Planning (grant number PXM2020_026283_000002). Data collection and dissemination for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI): the National Institutes of Health (grant number U01 AG024904), and the Department of Defense (award number W81XWH-12-2-0012). ADNI is funded by the National Institute of Aging and the National Institute of Biomedical Imaging and Bioengineering as well as through generous contributions from the following organizations: AbbVie, Alzheimer's Association, Alzheimer's Drug Discovery Foundation, Araclon Biotech, BioClinica Inc., Biogen, Bristol-Myers Squibb Company, CereSpir Inc., Eisai Inc., Elan Pharmaceuticals Inc., Eli Lilly and Company, EuroImmun, F. Hoffmann-La Roche Ltd. and its affiliated company Genentech Inc., Fujirebio, GE Healthcare, IXICO Ltd., Janssen Alzheimer Immunotherapy Research & Development LLC., Johnson & Johnson Pharmaceutical Research & Development LLC., Lumosity, Lundbeck, Merck & Co. Inc., Meso Scale Diagnostics LLC., NeuroRx Research, Neurotrack Technologies, Novartis Pharmaceuticals Corporation, Pfizer Inc., Piramal Imaging, Servier, Takeda Pharmaceutical Company, and Transition Therapeutics. The Canadian Institutes of Health Research are providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego, CA, USA. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California, CA, USA.

REFERENCES

1. Shiels PG, Steinvinkel P, Kooman JP, Mcguinness D. Circulating markers of ageing and allostatic load: a slow train coming. *Pract Lab Med.* (2017) 7:49–54. doi: 10.1016/j.plabm.2016.04.002
2. Khaw KT. Healthy aging. *BMJ.* (1997) 315:1090. doi: 10.1136/bmj.315.7115.1090
3. Rodrigue KM, Kennedy KM, Devous MD, Rieck JR, Hebrank AC, Diaz-Arrastia R, et al. β-Amyloid burden in healthy aging: regional distribution and cognitive consequences. *Neurology.* (2012) 78:387. doi: 10.1212/WNL.0b013e318245d295
4. Hardy JA, Higgins GA. Alzheimer's disease: the amyloid cascade hypothesis. *Science.* (1992) 256:184–5. doi: 10.1126/science.1566067
5. Armstrong RA. β-Amyloid (Aβ) deposition in elderly non-demented patients and patients with Alzheimer's disease. *Neurosci Lett.* (1994) 178:59–62. doi: 10.1016/0304-3940(94)90289-5
6. Aizenstein HJ, Nebes RD, Saxton JA, Price JC, Mathis CA, Tsopelas ND, et al. Frequent Amyloid Deposition Without Significant Cognitive Impairment Among the Elderly. *Arch Neurol.* (2008) 65:1509–17. doi: 10.1001/archneur.65.11.1509
7. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Trojanowski JQ. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol.* (2009) 65:403–13. doi: 10.1002/ana.21610
8. Sojkova J, Zhou Y, An Y, Kraut MA, Resnick SM. Longitudinal patterns of β-amyloid deposition in nondemented older adults. *Arch Neurol.* (2011) 68:644–9. doi: 10.1001/archneurol.2011.77
9. Price JL, Morris JC. Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. *Ann Neurol.* (1999) 45:358–68. doi: 10.1002/1531-8249(199903)45:3<358::AID-ANA12>3.0.CO;2-X
10. Hélène A, Hélène JG, Jean-Marc O, Le CN, Catherine H, Luc L, et al. The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study. *Brain.* (2005) 128(Pt 5):1093–101. doi: 10.1093/brain/awh451
11. Rowe CC, Ellis KA, Rimajova M, Bourgeat P, Villemagne VL. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. *Neurobiol Aging.* (2010) 31:1275–83. doi: 10.1016/j.neurobiolaging.2010.04.007
12. Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol.* (2013) 12:357–67. doi: 10.1016/S1474-4422(13)70044-9
13. Haroutunian V, Perl DP, Purohit DP, Marin D, Mohs RC. Regional distribution of neuritic plaques in the nondemented elderly and subjects with very mild Alzheimer disease. *Arch Neurol.* (1998) 55:1185–91. doi: 10.1001/archneur.55.9.1185
14. Wolf DS, Gearing M, Snowdon DA, Mori H, Markesbery WR, Mirra SS. Progression of regional neuropathology in Alzheimer disease and normal elderly: findings from the Nun study. *Alzheimer Dis Assoc Disord.* (1999) 13:226–31. doi: 10.1097/00002093-199910000-00009
15. Scheinin NM, Wikman K, Jula A, Perola M, Rinne JO. Cortical 11C-PIB Uptake is Associated with Age, APOE Genotype, and Gender in "Healthy Aging". *J Alzheimers Dis.* (2014) 41:193–202. doi: 10.3233/JAD-132783
16. Patterson BW, Elbert DL, Mawuenyega KG, Kasten T, Ovod V, Ma S, et al. Age and amyloid effects on human central nervous system amyloid-beta kinetics. *Ann Neurol.* (2015) 78:439–53. doi: 10.1002/ana.24454
17. Jack CR, Lowe VJ, Senjem ML, Weigand SD, Kemp BJ, Shiung MM, et al. 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. *Brain.* (2008) 131(Pt 3):665–80. doi: 10.1093/brain/awm336
18. Clark CM, Schneider JA, Bedell BJ, Beach TG, Skovronsky DM. Use of florbetapir-PET for imaging beta-amyloid pathology. *JAMA.* (2011) 305:275–83. doi: 10.1001/jama.2010.2008
19. Camus V, Payoux P, Barré L, Desgranges B, Voisin T, Tauber C, et al. Using PET with 18F-AV-45 (florbetapir) to quantify brain amyloid load in a clinical environment. *Eur J Nucl*

Med Mol Imaging. (2012) 39:621–31. doi: 10.1007/s00259-011-2021-8
20. Choi SR, Golding G, Zhuang Z, Zhang W, Lim N. Preclinical Properties of 18F-AV-45: a PET agent for Aβ plaques in the brain. *J Nucl Med.* (2009) 50:1887. doi: 10.2967/jnumed.109.065284
21. Lin KJ, Hsu WC, Hsiao IT, Wey SP, Kung MP. Whole-body biodistribution and brain PET imaging with [18F]AV-45, a novel amyloid imaging agent—a pilot study. *Nucl Med Biol.* (2010) 37:497–508. doi: 10.1016/j.nucmedbio.2010.02.003
22. Chen K, Roontiva A, Thiyyagura P, Lee W, Liu X, Ayutyanont N, et al. Improved power for characterizing longitudinal amyloid-β PET changes and evaluating amyloid-modifying treatments with a cerebral white matter reference region. *J Nucl Med.* (2015) 56:560. doi: 10.2967/jnumed.114.149732
23. Morel P. Gramm: grammar of graphics plotting in Matlab. *J Open Source Softw.* (2018) 3:568. doi: 10.21105/joss.00568
24. Woodard JL, Sugarman MA. Functional magnetic resonance imaging in aging and dementia: detection of age-related cognitive changes and prediction of cognitive decline. *Curr Top Behav Neurosci.* (2011) 10:113–36. doi: 10.1007/7854_2011_159
25. Long X, Zhang L, Jiang C, Zhang H. Age effect on structural hemispheric asymmetry revealed by multivariate model. *J Integration Technol*. (2016). Available online at: https://en.cnki.com.cn/Article_en/CJFDTotal-JCJI201603001.html
26. Brown JW, Jaffe J. Hypothesis on cerebral dominance. *Neuropsychologia.* (1975) 13:107–10. doi: 10.1016/0028-3932(75)90054-8
27. Albert MS. Geriatric neuropsychology. *J Consult Clin Psychol.* (1981) 49:835–50. doi: 10.1037/0022-006X.49.6.835
28. Henry E, Anton F, Ove A, Gunnar B, Emma L, Irina S, et al. Two-year follow-up of amyloid deposition in patients with Alzheimer's disease. *Brain.* (2006) 129(Pt 11):2856–66. doi: 10.1093/brain/awl281
29. Shin J, Lee SY, Kim SJ, Kim SH, Cho SJ, Kim YB. Voxel-based analysis of Alzheimer's disease PET imaging using a triplet of radiotracers: PIB, FDDNP, and FDG. *NeuroImage.* (2010) 52:488–96. doi: 10.1016/j.neuroimage.2010.04.013
30. Cho Y, Um YH, Kim TW, Seo HJ, Jeong JH, Hong SC, et al. A voxel wise analysis of cerebral beta amyloid retention in healthy controls and subjects with amnestic mild cognitive impairment and Alzheimer's Disease. *J Korean Geriatr Psychiatry.* (2016) 20:75–9. doi: 10.0000/jkgp.2016.20.2.75
31. Micheli C, Schepers IM, Ozker M, Yoshor D, Beauchamp MS, Rieger JW. Electrocorticography reveals continuous auditory and visual speech tracking in temporal and occipital cortex. *Eur J Neurosci.* (2018) 51:1364–76. doi: 10.1111/ejn.13992
32. Kondo HM, Kochiyama T. Normal aging slows spontaneous switching in auditory and visual bistability. *Neuroscience.* (2017) 389:152–60. doi: 10.1016/j.neuroscience.2017.04.040
33. Beer AL, Tina P, Georg M, Greenlee MW. Combined diffusion-weighted and functional magnetic resonance imaging reveals a temporal-occipital network involved in auditory-visual object processing. *Front Integr Neurosci.* (2013) 7:5. doi: 10.3389/fnint.2013.00005
34. Mrak RE, Griffin WST. Glia and their cytokines in progression of neurodegeneration. *Neurobiology of Aging.* (2005) 26:349–54. doi: 10.1016/j.neurobiolaging.2004.05.010
35. Sastre M, Walter J, Gentleman SM. *Interactions between APP secretases and inflammatory* mediators. *J Neuroinflammation.* (2008) 5:25. doi: 10.1186/1742-2094-5-25
36. Sheng JG, Mrak RE, Griffin WST. Enlarged and phagocytic, but not primed, interleukin-1α-immunoreactive microglia increase with age in normal human brain. *Acta Neuropathol (Berl).* (1998) 95:229–34. doi: 10.1007/s004010050792
37. Guillozet AL, Weintraub S, Mash DC, Mesulam MM. Neurofibrillary Tangles, Amyloid, and Memory in Aging and Mild Cognitive Impairment. *Archives of Neurology.* (2003) 60:729. doi: 10.1001/archneur.60.5.729

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 15 January 2021
doi: 10.3389/fmed.2020.605746

Texture Analysis of ^{18}F-FDG PET/CT for Differential Diagnosis Spinal Metastases

Xin Fan[1†], Han Zhang[1†], Yuzhen Yin[2†], Jiajia Zhang[1], Mengdie Yang[1], Shanshan Qin[1], Xiaoying Zhang[1*] and Fei Yu[1*]*

[1] Department of Nuclear Medicine, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China, [2] Shanghai Clinical College, Anhui Medical University, Shanghai, China

OPEN ACCESS

Edited by:
Xiaoli Lan, Huazhong University of Science and Technology, China

Reviewed by:
Salvatore Annunziata, Catholic University of the Sacred Heart, Italy
Susanne Asenbaum-Nan, Lower Landeskliniken-Holding, Austria

***Correspondence:**
Shanshan Qin
qinshanshan163@163.com
Xiaoying Zhang
yingzx2002@126.com
Fei Yu
yufei_021@163.com

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: *13 September 2020*
Accepted: *07 December 2020*
Published: *15 January 2021*

Citation:
Fan X, Zhang H, Yin Y, Zhang J, Yang M, Qin S, Zhang X and Yu F (2021) Texture Analysis of ^{18}F-FDG PET/CT for Differential Diagnosis Spinal Metastases. Front. Med. 7:605746. doi: 10.3389/fmed.2020.605746

Purpose: To evaluate the value of texture analysis for the differential diagnosis of spinal metastases and to improve the diagnostic performance of 2-deoxy-2-[fluorine-18]fluoro-D-glucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) for spinal metastases.

Methods: This retrospective analysis of patients who underwent PET/CT between December 2015 and January 2020 at Shanghai Tenth People's Hospital due to high FDG uptake lesions in the spine included 45 cases of spinal metastases and 44 cases of benign high FDG uptake lesions in the spine. The patients were randomly divided into a training group of 65 and a test group of 24. Seventy-two PET texture features were extracted from each lesion, and the Mann-Whitney *U*-test was used to screen the training set for texture parameters that differed between the two groups in the presence or absence of spinal metastases. Then, the diagnostic performance of the texture parameters was screened out by receiver operating characteristic (ROC) curve analysis. Texture parameters with higher area under the curve (AUC) values than maximum standardized uptake values (SUVmax) were selected to construct classification models using logistic regression, support vector machines, and decision trees. The probability output of the model with high classification accuracy in the training set was used to compare the diagnostic performance of the classification model and SUVmax using the ROC curve. For all patients with spinal metastases, survival analysis was performed using the Kaplan-Meier method and Cox regression.

Results: There were 51 texture parameters that differed meaningfully between benign and malignant lesions, of which four had higher AUC than SUVmax. The texture parameters were input to build a classification model using logistic regression, support vector machine, and decision tree. The accuracy of classification was 87.5, 83.34, and 75%, respectively. The accuracy of the manual diagnosis was 84.27%. Single-factor survival analysis using the Kaplan-Meier method showed that intensity was correlated with patient survival.

Conclusion: Partial texture features showed higher diagnostic value for spinal metastases than SUVmax. The machine learning part of the model combined with the texture parameters was more accurate than manual diagnosis. Therefore, texture analysis may be useful to assist in the diagnosis of spinal metastases.

Keywords: spinal metastases, texture analysis, PET/CT, diagnosis, machine learning

INTRODUCTION

The spine is the third most common site of metastatic disease after the lungs and liver, with ∼60–70% of patients with systemic cancer developing spinal metastases (1). Early and correct diagnosis of spinal metastases is helpful to guide clinical treatment, improve prognosis, and increase the survival rate. Computed tomography (CT) and magnetic resonance imaging (MRI) have their own advantages and disadvantages with regards to the diagnosis of spinal metastases. Two-deoxy-2-[fluorine-18]fluoro-D-glucose (^{18}F-FDG) imaging is more sensitive than conventional imaging examinations. It can facilitate early detection of the lesion by monitoring changes in glucose metabolism and can determine metastases of both the whole body skeletal system and soft tissues in a single examination. However, there is a certain degree of false-positive influence on the diagnosis, such as trauma-induced vertebral fractures, bone hyperplasia, and metabolic bone disease, which may interfere with the diagnosis.

The maximum standardized uptake value (SUVmax) is an intuitive quantitative measure of tissue ^{18}F-FDG uptake in current positron emission tomography/computed tomography (PET/CT) diagnostics. However, the currently established SUVmax diagnostic threshold has no clear criteria for spinal metastases. SUVmax is easy to use but does not fully reflect tumor size or tumor heterogeneity (2). Therefore, it is important to use texture analysis to extract more parameters in order to improve the diagnostic accuracy of metastatic lesions.

Texture analysis is a set of computational methods that extracts information regarding the relationship between adjacent pixels or voxels and assesses inhomogeneity, which can reflect the degree of benign or malignant properties and pathological features of the tissue (3, 4). Current applications are mainly based on CT and MRI, while lesser applications are applied to texture analysis using PET. Texture analysis using PET is more closely related to biological activity than texture parameters derived from CT and MRI. Furthermore, recent studies have shown that the grayscale texture variation characteristics of tumors in PET images can be used to evaluate the amount and unevenness of FDG uptake, which can be used to quantify the degree of tumor heterogeneity. For example, several studies have demonstrated that texture analysis can be used to find more powerful imaging biomarkers highly relevant to modern cancer therapy. Moreover, personalized treatment can be achieved through non-invasive molecular and genomic mapping of tumors (5–7).

In this study, spinal metastases or benign high ^{18}F-FDG uptake lesions were selected to analyze the diagnostic value of texture parameters derived from adjunctive PET/CT for spinal metastases. The purpose of this study was to initially verify whether texture analysis has a diagnostic value for spinal metastases and to screen out texture parameters that have better clinical guidance for the diagnosis of spinal metastases than those of traditional diagnostic methods.

MATERIALS AND METHODS

Study Population

This retrospective analysis was approved by the Ethics Committee of the Shanghai Tenth People's Hospital (SHSY-IEC-4.1/20-150/01) and registered with the Chinese Clinical Trials Registry (ChiCTR2000038089). We collected ^{18}F-FDG PET/CT image data of patients who had a positive spinal uptake of FDG and were admitted to the Shanghai Tenth People's Hospital between December 2015 and January 2020. Inclusion criteria were as following: complete electronic medical records or clear histopathological confirmation of spinal metastases, PET/CT image quality meeting diagnostic requirements, and basic clinical information available. Exclusion criteria were as following: primary spinal tumor, treated patients, tumor margins too difficult to delineate, incomplete clinical information, image quality not meeting diagnostic requirements, irregular spine, and inability to complete the follow-up. Patients with confirmed metastases all had not received antineoplastic therapy and either had undergone at least 6 months of follow-up or ended in death, moreover, patients with benign spinal uptake of FDG had no history of tumor.

Scanner and Acquisition Protocol

Image acquisition was performed using uMI 510 PET/CT. ^{18}F-FDG was manufactured by the Shanghai Xinke Pharmaceutical Company. The patient was fasted for more than 6 h before the examination, and the blood glucose was controlled <11.0 mmol/L, patients were administered ^{18}F-FDG at a dose of 0.10–0.15 mCi/Kg as the standard, and rested calmly for 60 min after intravenous injection of ^{18}F-FDG. The patient was placed in supine position during the scan, and the scanning range was from the skull base to the upper 1/3 of the femur. The PET images were reconstructed using the ordered subset maximum expected value iterative method (OS-EM), with image attenuation correction using CT scan data. The images were transferred to an Ulead workstation for frame-to-frame image alignment and fusion display.

Image Analysis

Data Grouping

The gold standard for the diagnosis of spinal metastases was clinical follow-up or pathologic confirmation of the lesion. The diagnosis was made by two nuclear medicine physicians with more than 10 years experiences in PET/CT diagnosis, without providing the patient's medical history or clinical data. The main criteria for manual diagnosis were osteolytic, osteogenic, or mixed bone lesions on PET/CT fusion images, partially accompanied by soft tissue mass formation, and abnormal increasing of FDG metabolism in the corresponding areas.

Extraction and Analysis of Texture Features

A nuclear medicine physician with more than 5 years of experience in PET/CT diagnosis sketched the high FDG metabolic lesions. A total of 6,408 heterogeneity indices were extracted from 72 PET texture features to evaluate the diagnostic value of PET texture analysis for spinal metastases. Six belonged to the co-occurrence matrix (C), 11 to the Voxel-alignment matrix, five to the neighborhood intensity-difference (NID), 11 to the intensity-size-difference zone matrix (ISZ), seven to the normalized co-occurrence (NC), 13 to the voxel statistics, two to the texture spectrum, three to the texture feature coding, nine to the texture feature coding co-occurrence, five to the neighborhood gray level dependence (NGLD) matrix. Detailed characterizations have been reported in previous studies (8).

Diagnostic Model Construction and Evaluation

We randomly divided the patients into a training group and a test group using the R 4.0.2 and set the seed number to 300. The construction of the machine learning model was based on the Python 6.3 platform. Moreover, the three machine learning models, logistic regression, support vector machine, and decision tree were trained and tested using the sklearn package. In both training group and testing group, the logistic regression model was compared with the diagnostic performance of SUVmax by plotting receiver operating characteristics (ROC) on the probability output of the training group.

Statistical Analysis

Analysis was performed using the SPSS 23.0 (IBM Statistics, New York, USA) and MedCalc Statistical Software version 15.2.2 (MedCalc Software bvba, Ostend, Belgium; http://www.medcalc.org; 2015) for analysis. Data conforming to a normal distribution are described as mean ± standard deviation (SD) of the overall distribution, otherwise as interquartile range (IQR). Data from the texture analysis were subjected to the Mann-Whitney *U*-test, and parameters with statistically significant differences were selected to plot the ROC curve and calculate the area under the curve (AUC). Spearman's test was used to analyze the correlation between texture parameters. DeLong's test was used to determine the difference between the ROC curves. The Kaplan-Meier method and Cox regression were used for single-factor and multi-factor survival analysis, respectively, and $P < 0.05$ was considered to be statistically significant.

Survival Analysis

The follow-up cutoff date was August 10, 2020; the endpoint event was patient death, and the follow-up period ranged from 3 to 40 months with an average follow-up time of 10.7 months. Single-factor and multi-factor survival analysis and mapping were performed using the R 4.0.2.

RESULT

Basic Patient Information

A total of 45 patients with confirmed spinal metastases and 44 patients with positive spinal uptake of FDG were analyzed, outlining a total of 89 lesions. After randomization into either the training or test group, the basic information and statistical differences between the two groups were compared, as shown in **Table 1**.

Distinction Between Tumor and Normal Group

After examining the normal distribution of the 72 texture features, we analyzed the differences between the two groups of texture parameters by Mann-Whitney *U*-test. Fifty-one texture parameters with different significance were selected and pre-processed. According to the heat map results, different colored partitions between the benign and malignant could be found (**Figure 1**).

For the texture parameters with statistically significant differences, we performed the ROC curve analysis. Finally, we took out the five parameters with diagnostic values better than SUVmax and performed correlation analysis on six parameters including SUVmax. Spearman's rank correlation coefficient was used to evaluate the correlation, and the results showed that SUV Variance was strongly correlated with SUV SD, so we eliminated SUV SD. The overall situation and distribution of the remaining five parameters in the training and testing groups are shown in **Table 2**. Finally, the optimal diagnostic threshold and its corresponding diagnostic sensitivity and specificity were found based on the Youden index (**Table 3**). The highest AUC value was the SUL peak, and the higher the value, the greater the degree of malignancy of the lesion, which is called a positive correlation, whereas the intensity value was inversely correlated with the degree of malignancy of the lesion.

Diagnostic Modeling and Performance Analysis

The jointly selected texture parameters were used to build a diagnostic model from the training group, which was constructed using three machine learning methods, logistic regression, decision tree, and support vector machine. The accuracy of classification in the test set was 87.5, 83.34, and 75%, respectively. The accuracy of the manual diagnosis was 84.27%. The combined model from logistic regression was used to plot the ROC curve with the probability output from the training group, and an AUC = 0.902 (95% CI = 0.803–0.962) was calculated. The DeLong test was used to compare the ROC curves of the combined model and SUVmax by logistic regression, and the results showed that the ROCs of the two groups were different ($P = 0.0345$), indicating

TABLE 1 | Clinical characteristics of patients in training and test groups.

	Training cohort ($n = 65$)			Test cohort ($n = 24$)				
	Tumor ($n = 33$)	Normal ($n = 32$)	P	Tumor ($n = 12$)	Normal ($n = 12$)	P	P_a	P_b
Age (years)			0.106			0.181	0.100	0.399
Mean ± SD	64.67 ± 8.324	59.44 ± 16.274		60.25 ± 5.956	63.67 ± 8.669			
Range	39–80	27–84		52–73	52–81			
Sex			0.015			0.041	0.787	0.323
Male	26	16		9	4			
Female	7	16		3	8			

The P_a was derived from the Student's t or chi-square test of tumor groups between the training and test cohorts and the P_b was derived from that of normal groups between the training and test cohorts.

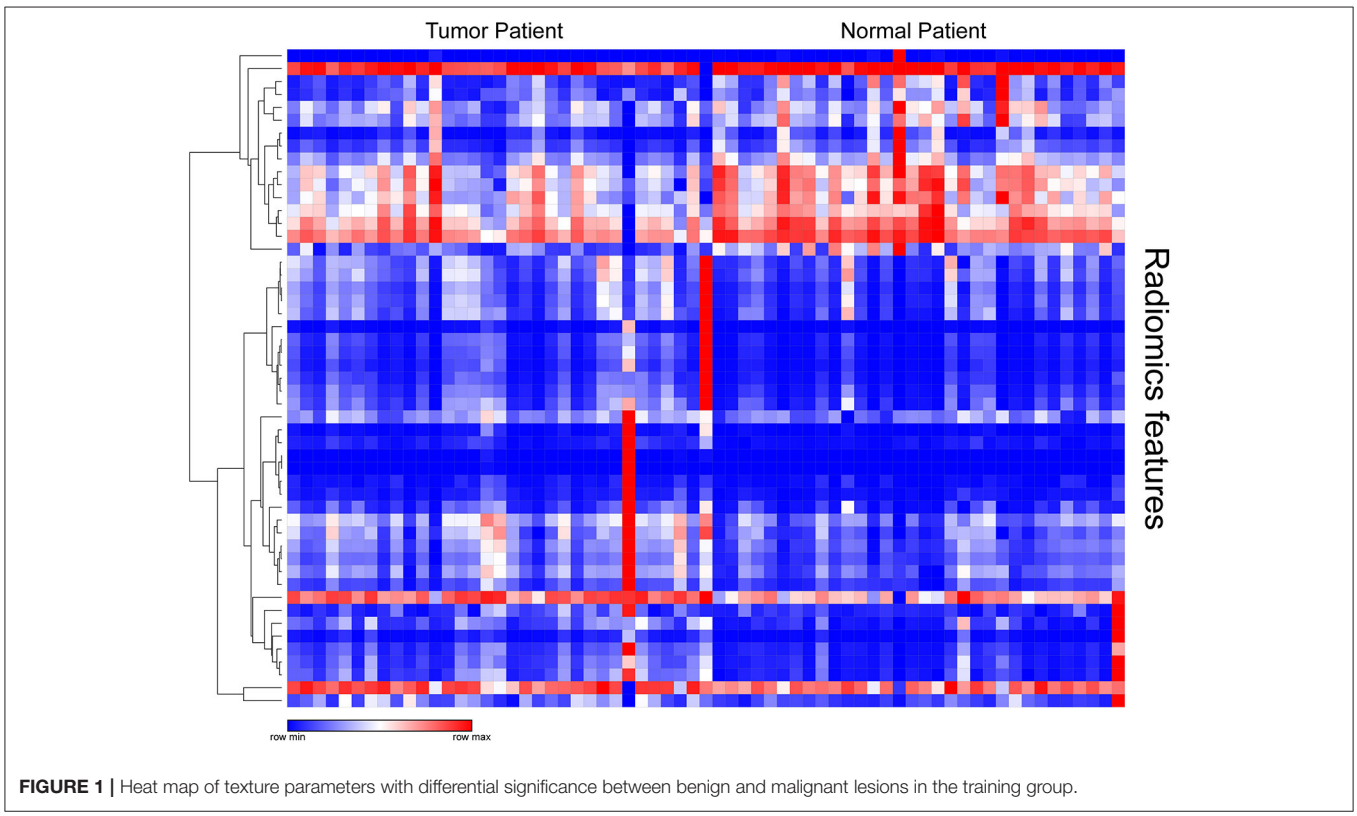


FIGURE 1 | Heat map of texture parameters with differential significance between benign and malignant lesions in the training group.

that the diagnostic model constructed jointly with the texture analysis parameters had a better diagnostic value than SUVmax (**Figure 2**) (9).

Survival Analysis

The 45 patients with spinal metastases were followed for 3–40 months, with a median follow-up time of 8 months. By the last follow-up, 37 of the 45 patients with spinal metastases had died, and the median survival time for patients with metastatic spinal tumors was 8 months. The 1- and 2-year survival rates of patients were 48.9 ± 7.5% and 19.4 ± 6.0%, respectively (**Figure 3**). We performed a single-factor survival analysis of the five texture parameters and survival by classifying the five texture parameters into low and high groups based on the median of the texture parameters as the criterion, and the results showed that patients in the low intensity group had shorter survival ($P = 0.041$) (**Figure 3**). The multi-factor survival analysis showed that none of the five texture parameters and survival outcomes were statistically significant ($P > 0.05$).

DISCUSSION

Our study demonstrates the significance of PET/CT texture analysis for the analysis of spinal positive FDG uptake lesions for the identification and diagnosis of spinal metastases. We selected five texture parameters of diagnostic significance, four

TABLE 2 | Differences in the overall distribution of parameters with good diagnostic value between the benign and malignant groups.

	Training cohort			Test cohort				
	Tumor ($n = 33$) Median (IQR)	Normal ($n = 32$) Median (IQR)	*P*	Tumor ($n = 12$) Median (IQR)	Normal ($n = 12$) Median (IQR)	*P*	*Pa*	*Pb*
SUL peak	5.97 (3.98, 8.10)	3.19 (2.70, 4.07)	0.001	6.82 (5.03, 8.61)	2.89 (2.49, 3.16)	0.001	0.603	0.095
Correlation	0.81 (0.74, 0.84)	0.69 (0.63, 0.74)	0.001	0.82 (0.78, 0.83)	0.62 (0.56, 069)	0.001	0.676	0.040
Intensity*	255.69 (247.10, 263.64)	278.10 (267.26, 289.10)	0.001	246.35 (234.49, 264.29)	258.00 (272.93, 297.01)	0.001	0.291	0.118
SUV Variance	2.42 (0.75, 4.61)	0.31 (0.21, 0.80)	0.001	3.23 (1.15, 5.89)	0.29 (0.22, 0.38)	0.001	0.568	0.328
Maximum SUV	10.06 (6.37, 12.69)	5.61 (4.50, 6.82)	0.001	11.64 (7.18, 14.31)	4.97 (4.45, 5.71)	0.001	0.409	0.131

The Pa was derived from the Mann-Whitney U-test of tumor groups between the training and test cohorts and the Pb was derived from that of normal groups between the training and test cohorts.

TABLE 3 | AUC values of five texture parameters and associated diagnostic performance parameters incorporated into the model.

Texture feature	AUC	CI 95%	Cut-off	Se (%)	Sp (%)
SUL peak	0.831	0.728, 0.935	3.582	90.9	65.6
Correlation	0.827	0.719, 0.935	0.738	84.8	75.0
Intensity*	0.820	0.715, 0.925	263.577	75.8	81.2
SUV variance	0.817	0.709, 0.926	0.649	78.8	75.0
Maximum SUV	0.806	0.696, 0.915	7.40	69.7	84.4

with a better diagnostic performance than current diagnostic modalities, and established a diagnostic model with a greater diagnostic performance than manual diagnosis. It shows the texture analysis can assist in the differential diagnosis spinal metastases and may have better diagnostic value than current methods.

There are two highlights of our study. Firstly, all the patients enrolled were patients with spinal metastases confirmed by definite pathological findings or follow-up results, and complete follow-up data on survival time were available. Secondly, to the best of our knowledge, this is the first time report that PET texture analysis has been used in the diagnosis of spinal metastases.

Radiomics is achieved progressively through segmentation of lesions, feature data extraction, database creation, and analysis of individualized data, while texture analysis is a class of feature data extraction with objective descriptive features. Most of the traditional imaging distributions of lesions are subjectively described, and correlating the results of texture analysis with the subjectively described features can make the diagnostic conclusions more convincing. ^{18}F-FDG PET/CT texture analysis can provide more detail regarding tumor spatial information and tumor heterogeneity than clinically used parameters such as SUVmax, metabolic tumor volume (MTV), and total lesion glycolysis (TLG). Although PET/CT is superior to bone scan and CT for the diagnosis of spinal metastases (10), it is currently still prone to false positives for the diagnosis of spinal metastases. The SUVmax of Schmorl's nodes is similar to that of spinal metastases (11). High FDG uptake also occurs at different times after a benign fracture, so false-positive results may occur when performing ^{18}F-FDG PET/CT imaging to assess metastases, although different uptake modalities and clinical correlations usually allow accurate differentiation of fractures from skeletal metastases (12). In addition, spinal metastases need to be differentiated from discontinuous spinal tuberculosis and spinal degenerative diseases (13, 14).

The robustness of the texture analysis software we use to measure texture values has been verified previously (8). PET/CT texture analysis has been shown to be effective in diagnosing and predicting prognosis in a variety of diseases. Bianconi et al. found a significant correlation between PET features, CT features, and histological type in non-small cell lung cancer (NSCLC) and texture analysis shows the potential for differentiating histological types in NSCLC (15). Feliciani et al. found that texture analysis of ^{18}F-FDG PET has predictive value for the effectiveness of treatment of primary head and neck squamous cell carcinoma (HNSCC) treated with concurrent chemoradiotherapy (16). Lovinfosse et al. analyzed the SUVmax and mean standard uptake value (SUVmean), MTV, TLG, and 13 global, local, and regional texture features of 63 NSCLC patients undergoing stereotactic body radiotherapy (SBRT) who underwent ^{18}F-FDG PET/CT prior to treatment. They found that differences in texture features measured at baseline ^{18}F-FDG PET/CT appeared to be a strong independent predictor of prognosis in SBRT-treated NSCLC patients (17). Xu et al. found that the texture fractionation method of PET is useful for the differential diagnosis of benign and malignant bone and soft tissue lesions, in which the texture parameters coarseness and entropy have better diagnostic performance than SUV (18).

Gao et al. extracted the texture parameters from PET/CT images to build a support vector machine that can identify benign and malignant mediastinal lymph nodes in NSCLC patients (19). Oh et al. successfully evaluated the efficacy and survival of 70 patients with hypopharyngeal cancer after radiotherapy and chemotherapy by roughness in pre-treatment texture parameters (20). Pyka et al. found that texture analysis on PET images not only allowed some assessment of the local recurrence of NSCLC patients after radiotherapy, but also predicted their long-term survival (21).

Most tumors develop spinal metastases earlier; therefore, the characteristics of spinal metastases correlate with the overall survival prognosis of patients. The survival time of patients with

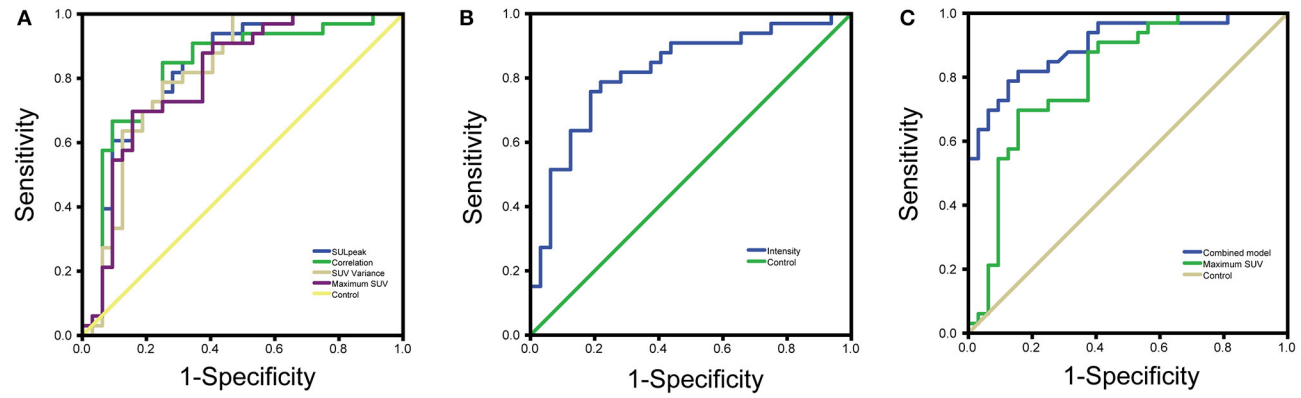


FIGURE 2 | (A) ROC curves for four parameters with good diagnostic value **(B)** ROC curves for texture analysis parameters with inverse proportion of size and degree of malignancy **(C)** ROC curves for combined model diagnostic parameters vs. maximum SUV.

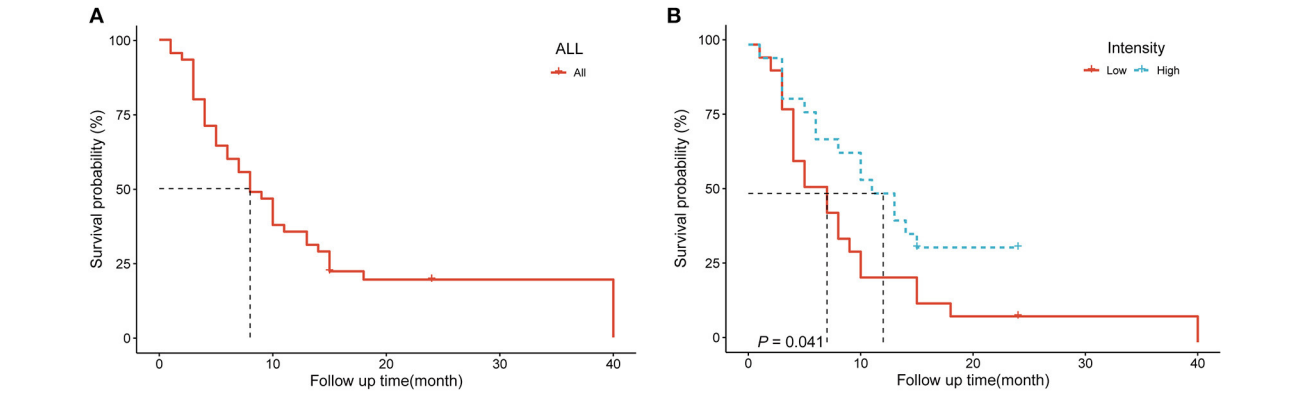


FIGURE 3 | (A) Survival analysis curves for 45 patients with spinal metastases **(B)** Survival analysis of patients in different intensity groups.

metastases is related to a greater degree with the primary tumor and the location and number of metastases. Moreover, the only set of texture parameters with low intensities was found in the single-factor analysis. Patients in low intensity group had a worse prognosis, and multi-factor Cox regression analysis remained insignificant. It is possible that this texture parameter can be used to predict patient prognosis, but a large, prospective study is needed for further validation. The pathological diagnosis of spinal metastases in clinical practice is limited by inadequate access to tissue; it is more traumatic and unacceptable to most patients. Currently, there are three main approaches to the treatment of spinal metastases, chemotherapy, radiation therapy, and surgery. The goals of both medical and surgical treatment of metastases are to maximize the improvement in quality of life. Once a diagnosis of metastasis is established, the role of surgery or surgery in combination with other treatments can relieve pain, improve or maintain neurological function, and restore the structural integrity of the spine (22, 23).

Limitations

There are certain limitations to our study. Only the relationship between texture analysis and character of the disease was analyzed, which can be followed by further refinement of the molecular type of the disease and the classification of the serologic examination indexes to evaluate the value of texture analysis. In addition, this is a single-center study, and the data from different centers may have some influence on the stability of texture analysis due to the different methods of PET image reconstruction, which can be used to study large samples of texture analysis data according to different machines and diseases to explore the stability value of texture analysis.

CONCLUSIONS

We evaluated the feasibility of using texture analysis for the differential diagnosis of spinal metastases by analyzing and quantifying the overall distribution and heterogeneity of high FDG uptake spinal lesions, which can be effective in the diagnosis of spinal metastases. By carefully and objectively selecting out the PET texture parameters and deriving the optimal threshold to diagnose spinal metastases, certain texture features showed better diagnostic value than SUVmax. Thus, texture analysis in ^{18}F-FDG PET/CT images

may play a role in the differential diagnosis of spinal metastases, which provide more accurate and comprehensive guidance to clinical treatment.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of Shanghai Tenth People's Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

XF, HZ, YY, SQ, XZ, and FY: conception and design. XF: acquisition, statistical analysis, or interpretation of the data. All authors: drafting of the manuscript, reviewed, and approved the final version of the manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (No. 2016YFC0104303).

ACKNOWLEDGMENTS

We thank the patients who were included in this research project and the staff who developed the CGITA.

REFERENCES

1. Shah LM, Salzman KL. Imaging of spinal metastatic disease. *Int J Surg Oncol.* (2011) 2011:769753. doi: 10.1155/2011/769753
2. Takeda K, Takanami K, Shirata Y, Yamamoto T, Takahashi N, Ito K, et al. Clinical utility of texture analysis of 18F-FDG PET/CT in patients with Stage I lung cancer treated with stereotactic body radiotherapy. *J Radiation Res.* (2017) 58:862–9. doi: 10.1093/jrr/rrx050
3. Orlhac F, Thézé B, Soussan M, Boisgard R, Buvat I. Multiscale texture analysis: from ^{18}F-FDG PET images to histologic images. *J. Nucl. Med.* (2016) 57:1823–8. doi: 10.2967/jnumed.116.173708
4. Orlhac F, Soussan M, Maisonobe JA, Garcia CA, Vanderlinden B, Buvat I. Tumor texture analysis in ^{18}F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med.* (2014) 55:414–22. doi: 10.2967/jnumed.113.129858
5. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present… any future? *Eur J Nucl Med Mol Imaging.* (2017) 44:151–65. doi: 10.1007/s00259-016-3427-0
6. Lee H, Lee DE, Park S, Kim TS, Jung SY, Lee S, et al. Predicting response to neoadjuvant chemotherapy in patients with breast cancer: combined statistical modeling using clinicopathological factors and FDG PET/CT texture parameters. *Clin Nucl Med.* (2019) 44:21–9. doi: 10.1097/RLU.0000000000002348
7. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in ^{18}F-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging.* (2013) 40:133–40. doi: 10.1007/s00259-012-2247-0
8. Fang YH, Lin CY, Shih MJ, Wang HM, Ho TY, Liao CT, et al. Development and evaluation of an open-source software package "CGITA" for quantifying tumor heterogeneity with molecular images. *BioMed Res Int.* (2014) 2014:248505. doi: 10.1155/2014/248505
9. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* (1988) 44:837–45. doi: 10.2307/2531595
10. Uchida K, Nakajima H, Miyazaki T, Tsuchida T, Hirai T, Sugita D, et al. (18)F-FDG PET/CT for diagnosis of osteosclerotic and osteolytic vertebral metastatic lesions: comparison with bone scintigraphy. *Asian Spine J.* (2013) 7:96–103. doi: 10.4184/asj.2013.7.2.96
11. Wang Z, Ma D, Yang J. ^{18}F-FDG PET/CT can differentiate vertebral metastases from Schmorl's nodes by distribution characteristics of the ^{18}F-FDG. *Hellenic J Nucl Med.* (2016) 19:241–4. doi: 10.1967/s002449910406
12. Shon IH, Fogelman I. F-18 FDG positron emission tomography and benign fractures. *Clin Nucl Med.* (2003) 28:171–5. doi: 10.1097/01.RLU.0000053508.98025.01
13. Gao X, Ye XD, Yuan Z, Yang CS. Non-contiguous spinal tuberculosis with a previous presumptive diagnosis of lung cancer spinal metastases. *Eur J Cardio Thoracic Surg.* (2014) 45:e178. doi: 10.1093/ejcts/ezu047
14. Rosen RS, Fayad L, Wahl RL. Increased ^{18}F-FDG uptake in degenerative disease of the spine: characterization with ^{18}F-FDG PET/CT. *J Nucl Med.* (2006) 47:1274–80. Available online at: https://jnm.snmjournals.org/content/47/8/1274.long
15. Bianconi F, Palumbo I, Fravolini ML, Chiari R, Minestrini M, Brunese L, et al. Texture analysis on [(18)F]FDG PET/CT in non-small-cell lung cancer: correlations between PET features, CT features, and histological types. *Mol Imaging Biol.* (2019) 21:1200–9. doi: 10.1007/s11307-019-01336-3
16. Feliciani G, Fioroni F, Grassi E, Bertolini M, Rosca A, Timon G, et al. Radiomic profiling of head and neck cancer: (18)F-FDG PET texture analysis as predictor of patient survival. *Contrast Media Mol Imaging.* (2018) 2018:3574310. doi: 10.1155/2018/3574310
17. Lovinfosse P, Janvary ZL, Coucke P, Jodogne S, Bernard C, Hatt M, et al. FDG PET/CT texture analysis for predicting the outcome of lung cancer treated by stereotactic body radiation therapy. *Eur J Nucl Med Mol Imaging.* (2016) 43:1453–60. doi: 10.1007/s00259-016-3314-8
18. Xu R, Kido S, Suga K, Hirano Y, Tachibana R, Muramatsu K, et al. Texture analysis on (18)F-FDG PET/CT images to differentiate malignant and benign bone and soft-tissue lesions. *Ann Nucl Med.* (2014) 28:926–35. doi: 10.1007/s12149-014-0895-9
19. Gao X, Chu C, Li Y, Lu P, Wang W, Liu W, et al. The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from (18)F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer. *Eur J Radiol.* (2015) 84:312–7. doi: 10.1016/j.ejrad.2014.11.006
20. Oh JS, Kang BC, Roh JL, Kim JS, Cho KJ, Lee SW, et al. Intratumor textural heterogeneity on pretreatment (18)F-FDG PET images predicts response and survival after chemoradiotherapy for hypopharyngeal cancer. *Ann Surg Oncol.* (2015) 22:2746–54. doi: 10.1245/s10434-014-4284-3

21. Pyka T, Bundschuh RA, Andratschke N, Mayer B, Specht HM, Papp L, et al. Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy. *Radiation oncology.* (2015) 10:100. doi: 10.1186/s13014-015-0407-7
22. Bartels RH, van der Linden YM, van der Graaf WT. Spinal extradural metastasis: review of current treatment options. *CA Cancer J Clin.* (2008) 58:245–59. doi: 10.3322/CA.2007.0016
23. Spratt DE, Beeler WH, de Moraes FY, Rhines LD, Gemmete JJ, Chaudhary N, et al. An integrated multidisciplinary algorithm for the management of spinal metastases: an International Spine Oncology Consortium report. *Lancet Oncol.* (2017) 18:e720–30. doi: 10.1016/S1470-2045(17)30612-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 18 January 2021
doi: 10.3389/fmed.2020.621204

Use of a Sparse-Response Deep Belief Network and Extreme Learning Machine to Discriminate Alzheimer's Disease, Mild Cognitive Impairment, and Normal Controls Based on Amyloid PET/MRI Images

Ping Zhou, Shuqing Jiang, Lun Yu, Yabo Feng, Chuxin Chen, Fang Li, Yang Liu and Zhongxiong Huang on behalf of the Alzheimer's Disease Neuroimaging Initiative*

Department of PET-CT Center, Chenzhou No.1 People's Hospital, Chenzhou, China

OPEN ACCESS

Edited by:
Chuantao Zuo,
Fudan University, China

Reviewed by:
Woon-Man Kung,
Chinese Culture University, Taiwan
Juanjuan Jiang,
Shanghai University, China

***Correspondence:**
Zhongxiong Huang
hzx1968@yeah.net

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: *25 October 2020*
Accepted: *26 November 2020*
Published: *18 January 2021*

Citation:
Zhou P, Jiang S, Yu L, Feng Y, Chen C, Li F, Liu Y and Huang Z (2021) Use of a Sparse-Response Deep Belief Network and Extreme Learning Machine to Discriminate Alzheimer's Disease, Mild Cognitive Impairment, and Normal Controls Based on Amyloid PET/MRI Images. Front. Med. 7:621204. doi: 10.3389/fmed.2020.621204

In recent years, interest has grown in using computer-aided diagnosis (CAD) for Alzheimer's disease (AD) and its prodromal stage, mild cognitive impairment (MCI). However, existing CAD technologies often overfit data and have poor generalizability. In this study, we proposed a sparse-response deep belief network (SR-DBN) model based on rate distortion (RD) theory and an extreme learning machine (ELM) model to distinguish AD, MCI, and normal controls (NC). We used [^{18}F]-AV45 positron emission computed tomography (PET) and magnetic resonance imaging (MRI) images from 340 subjects enrolled in the ADNI database, including 116 AD, 82 MCI, and 142 NC subjects. The model was evaluated using five-fold cross-validation. In the whole model, fast principal component analysis (PCA) served as a dimension reduction algorithm. An SR-DBN extracted features from the images, and an ELM obtained the classification. Furthermore, to evaluate the effectiveness of our method, we performed comparative trials. In contrast experiment 1, the ELM was replaced by a support vector machine (SVM). Contrast experiment 2 adopted DBN without sparsity. Contrast experiment 3 consisted of fast PCA and an ELM. Contrast experiment 4 used a classic convolutional neural network (CNN) to classify AD. Accuracy, sensitivity, specificity, and area under the curve (AUC) were examined to validate the results. Our model achieved 91.68% accuracy, 95.47% sensitivity, 86.68% specificity, and an AUC of 0.87 separating between AD and NC groups; 87.25% accuracy, 79.74% sensitivity, 91.58% specificity, and an AUC of 0.79 separating MCI and NC groups; and 80.35% accuracy, 85.65% sensitivity, 72.98% specificity, and an AUC of 0.71 separating AD and MCI groups, which gave better classification than other models assessed.

Keywords: computer-aided diagnosis, Alzheimer's disease, mild cognitive impairment, sparse-response deep belief network, extreme learning machine

INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease characterized by cognitive dysfunction and associated with advanced age. Because there are currently no therapies that can reverse the course of AD, it is important to diagnose AD and its prodromal stage, mild cognitive impairment (MCI) as early as possible (1).

In recent years, neuroimaging techniques have been shown to be effective tools for the diagnosis of AD. Magnetic resonance imaging (MRI) and positron emission tomography (PET) are two common neuroimaging methods. For example, Hua et al. proposed a powerful tool to monitor structural atrophy in incipient stages of AD using MR images (2). Mosconi et al. demonstrated that PET scans may provide objective and sensitive support to clinical diagnosis in early dementia (3). In addition, deep learning methods have shown great promise for image analysis and disease prediction. For instance, Hu et al. utilized a targeted autoencoder network to classify functional connectivity matrices across brain regions, which was able to distinguish MCI from NC with 87.5% accuracy (4). Liu et al. designed a deep learning architecture to more accurately differentiate AD, MCI, and normal controls (NC). The architecture, including stacked autoencoders and a softmax output layer, achieved 87.76% accuracy, 88.57% sensitivity, and 87.22% specificity distinguishing AD from NC and exhibited 76.92% accuracy, 74.29% sensitivity, and 78.13% specificity distinguishing MCI from NC (5). In addition, a few of deep learning studies based on PET/MRI could also be observed (6, 7).

However, the methods mentioned above had some disadvantages. For instance, gradient diffusion and gradient explosion may emerge with deepening of the autoencoder stack depth, resulting in decreased classification accuracy. To mitigate this limitation, we proposed a sparse-response deep belief network (SR-DBN) based on the rate distortion (RD) theory model. Our SR-DBN used the contrastive divergence algorithm to maximize the retention of data distribution, in case gradient diffusion and gradient explosion became factors. In addition, the SR-DBN model included sparsity. Compared to DBN models without sparsity, sparse representations allow changing the significant bits for each example in a fixed-size representation, which are more efficient from the point of view of information theory (8). Subsequently, we used an extreme learning machine (ELM) as a classifier to get the performance of the classification. Meanwhile, to evaluate the effectiveness of our method, we compared our model with other models.

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's disease neuroimaging initiative; AUC, area under curve; CAD, computer-aided diagnosis; CNN, convolutional neural network; ELM, extreme learning machine; GM, gray matter; MCI, mild cognitive impairment; MRI, magnetic resonance imaging; NC, normal controls; PCA, principal components analysis; PET, positron emission computed tomography; RBM, restricted Boltzmann machine; RD, rate distortion; SR-DBN, sparse-response deep belief network; SVM, support vector machine; WM, white matter.

MODEL DESIGN

Model Framework

As shown in **Figure 1**, the framework of the model consists of four parts: (1) original image data underwent standard preprocessing; (2) data dimensionality was reduced using fast principal component analysis (PCA); (3) features were extracted by three SR-DBNs based on rate distortion theory; and (4) processed data were classified by the ELM.

Mathematical Fundamentals of the Proposed Model

SR-DBN Model Based on RD Theory

Restricted Boltzmann machines (RBM) are neural perceptrons composed of a visible layer and a hidden layer. Several RBMs can form a DBN. Similar to the structure of a DBN, the SR-DBN also consists of several sparse-response restricted Boltzmann machines (SR-RBMs). In the model, the Kullback–Leibler divergence $KL(p^0 \| p_\theta^\infty)$ (9) between the original data's distribution p^0 and the equilibrium distribution p_θ^∞ defined by RBM served as a distortion function. Considering the RD theory, we can deduce the following formulation:

$$\min_{w_{ij},c_i,b_j} KLp^0 \| p_\theta^\infty + \lambda \sum_{l=1}^{m} \left\| p\left(\mathbf{h}^{(\mathbf{l})} | \mathbf{v}^{(\mathbf{l})}\right) \right\|_1 \tag{1}$$

where $\sum_{l=1}^{m} \left\| p\left(\mathbf{h}^{(\mathbf{l})} | \mathbf{v}^{(\mathbf{l})}\right) \right\|_1$ denotes the sparseness of representation and λ is a regularization parameter. Then we replaced Kullback–Leibler divergence with $p^0 \| p_\theta^\infty - p_\theta^1 \| p_\theta^\infty$ to simplify calculations (10). Suppose w is the weight matrix of RBM, b is the bias vector of the input layer, and c is the bias vector of the output layer, giving the updated rules below:

$$\begin{cases} w_{ij} := w_{ij} + \epsilon \left(\langle v_i h_j \rangle_{p^0} - \langle v_i h_j \rangle_{p_\theta^1} \right) \\ c_i := c_i + \epsilon \left(\langle v_i \rangle_{p^0} - \langle v_i \rangle_{p_\theta^1} \right) \\ b_j = b_j + \epsilon \left(\langle h_j \rangle_{p^0} - \langle h_j \rangle_{p_\theta^1} \right) \end{cases} \tag{2}$$

where ϵ denotes a learning rate. Additionally, we added another update with the gradient of the regularization term in each iteration. The term is as follows:

$$\begin{aligned} -\frac{\partial}{\partial w_{ij}} \sum_{l=1}^{m} \left\| p\left(\mathbf{h}^{(\mathbf{l})} | \mathbf{v}^{(\mathbf{l})}\right) \right\|_1 &= -\sum_{l=1}^{m} \frac{\partial}{\partial w_{ij}} \text{sigmoid}\left(b_j + \sum_i w_{ij} v_i^{(l)} \right) \\ &= -\sum_{l=1}^{m} p_j^{(l)} \left(1 - p_j^{(l)}\right) v_i^{(l)} \end{aligned} \tag{3}$$

$$\begin{aligned} -\frac{\partial}{\partial b_j} \sum_{l=1}^{m} \left\| p\left(\mathbf{h}^{(\mathbf{l})} | \mathbf{v}^{(\mathbf{l})}\right) \right\|_1 &= -\sum_{l=1}^{m} \frac{\partial}{\partial w_{ij}} \text{sigmoid}\left(b_j + \sum_i b_j v_i^{(l)} \right) \\ &= -\sum_{l=1}^{m} p_j^{(l)} \left(1 - p_j^{(l)}\right) \end{aligned} \tag{4}$$

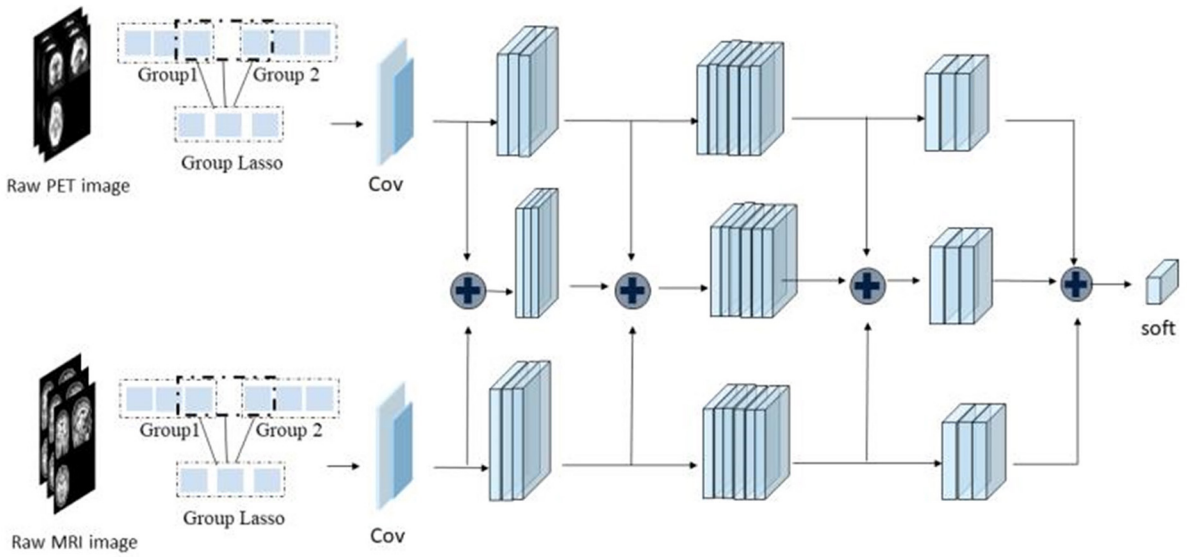


FIGURE 1 | The framework of this study.

where $p_j^{(l)} = \text{sigmoid}\left(\sum_i v_i^{(l)} w_{ij} + b_j\right)$, and sigmoid(.) represents the sigmoid function.

In this study, we employed one input layer, three hidden layers and one output layer.

ELM Model for Classification

An ELM is a neural network algorithm for a single hidden layer feedforward neural network. Its input weights and hidn node bias are generated randomly within a given range. The only optimal solution can be obtained by setting the number of hidden layer neurons (11). When the input weights and hidden layer bias are determined randomly, the output matrix of the hidden layer, H is also determined (12):

$$\hat{\beta} = H^{+}T \tag{5}$$

where H^{+} is the Moore–Penrose pseudoinverse matrix of H and the T notes the expected output.

MATERIALS AND METHODS

Materials

The data used in this study were access through the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database. ADNI is a consortium study initiated in 2004 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations (13). For additional information about ADNI, please see www.adni-info.org.

TABLE 1 | The clinical data of three cohorts.

	AD (n = 116)	MCI (n = 82)	NC (n = 142)
Gender (M/F)	68/48	50/32	52/90
Ages (years)	75.37 ± 5.61	75.6 ± 5.87	73.86 ± 7.06
MMSE	21.63 ± 3.97	23.8 ± 5.98	28.92 ± 1.27
MOCA	15.89 ± 5.85	18.8 ± 6.49	26.01 ± 2.76

In this study, we selected AV45 PET and structural MRI images of 340 subjects enrolled in ADNI, including 116 AD, 82 MCI, and 142 NC subjects. The clinical data for each of these diagnostic groups is shown in **Table 1**.

Image Preprocessing

MRI data were acquired on multiple 3T MRI scanners using scanner-specific T1-weighted sagittal 3D MPRAGE sequences. In order to increase signal uniformity across the multicenter scanner platforms, original MPRAGE acquisitions underwent standardized image preprocessing steps. The current study implemented the following steps: (1) segmentation of the images into gray matter (GM), white matter (WM) and cerebrospinal fluid (14), of which gray matter and white matter were used for further analysis; (2) normalization of all GM and WM images into Montreal Neurological Institute space; and (3) spatial smoothing using a Gaussian kernel of 4 mm^3.

[18F]-AV45 PET data were acquired on multiple instruments of varying resolutions and following different platform-specific acquisition protocols. Similar to the MRI data, ADNI PET data underwent standardized image preprocessing steps aimed at increasing data uniformity across the multicenter acquisitions (15). The preprocessing steps included realignment, spatial

normalization to MNI space, and smoothing using a 7-mm^3 Gaussian kernel. We performed a voxel-based partial volume effects correction of the normalized functional image using the Müller-Gärtner method. Lastly, the partial volume effect-corrected functional image was smoothed to reduce noise and improve image quality using an isotropic Gaussian smoothing kernel with a full width at half maximum setting of 7 mm^3. The image was scaled up to obtain a standard uptake value rate map of the entire cerebellum.

Both MRI and ^{18}F-AV45 PET images were preprocessed using statistical parametric mapping software (SPM12, https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) on Matlab 2016b.

Dimension Reduction and Feature Extraction

Fast PCA was used to describe the data with a small number of linearly independent features under the principle of ensuring the minimum loss of data information.

In the study, the SR-DBN model undertook feature extraction. Compared with DBN, the SR-DBN is more efficient from the perspective of information theory, which allows changing the effective number of bits per example in a fixed size representation (8).

The SR-DBN model used in the study was made up of multiple basic SR-RBMs with the same numbers of nodes. The output of each SR-RBM was the input of the next basic SR-RBM at successive levels. In the last layer of the SR-DBN model, a back propagation network was set, receiving the output feature vector of SR-RBM as learned features, and adopting a gradient descent algorithm to fine-tune the weight of the whole network, thereby coordinating and optimizing the parameters of the whole SR-DBN.

Classification & Comparative Experiments

Three kinds of images were used as input: [^{18}F]-AV45 PET, and GM and WM segmentations from the MRI. Correspondingly, we used three SR-DBNs to extract features. Following feature extraction, the ELM classified the three diagnostic groups. After obtaining the predicted labels, the accuracy, sensitivity, specificity, and area under curve (AUC) were calculated to evaluate the practicability of the model.

The model was evaluated using five-fold cross-validation, repeated 200 times. In the case of "lucky trails," we randomly sampled the training and testing instances from each class to ensure they had similar distributions as the original dataset. The entire network was trained and fine-tuned with 80% of the data and then tested with the remaining 20% of the samples in each validation trial.

To evaluate the effectiveness of our method, we performed several comparative trials. In contrast experiment 1, ELM was replaced by a support vector machine (SVM). Contrast experiment 2 utilized DBN without sparsity. Contrast experiment 3 consisted of fast PCA and ELM. Contrast experiment 4 used a classic convolutional neural network (CNN) to classify AD, MCI, and NC. The experimental platform is based on Matlab 2016b.

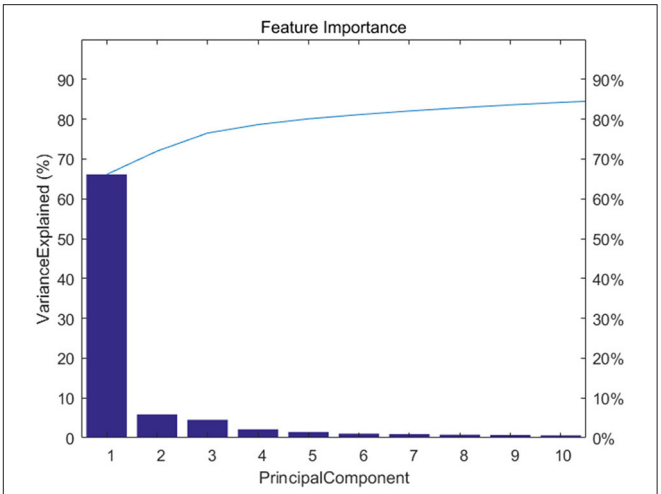


FIGURE 2 | The feature importance of fast PCA.

RESULTS

Results of Dimension Reduction

Figure 2 shows feature importance. We extracted the top 20 features which represent 90% information of the original data.

Results of Classification

The classification and comparative results are shown in **Table 2** and **Figure 3**. In the classification of AD and NC, our model achieved 91.68% accuracy, 95.47% sensitivity, 86.68% specificity, and an AUC of 0.87. In the classification between MCI and NC, the model achieved 87.25% accuracy, 79.74% sensitivity, 91.58% specificity, and an AUC of 0.79. When separating between AD and MCI, the model achieved 80.35% accuracy, 85.65% sensitivity, 72.98% specificity, and an AUC of 0.71. Moreover, the time cost for image processing and classification in our proposed method and four compared methods were 36.2 s, 37.4 s, 491.7 s, 25 s, and 1,386.5 s. This result means that our method is faster than the classical CNN model, and similar to machine learning models.

Table 3 shows the comparative results of our model and results from the literature, including Hu's model, Liu's model, and Suk's model (4, 5, 16). Specifically, Hu's model used a single image modality (MRI) and Liu's model used both MRI and PET. Our proposed model achieved the best classification result of all models compared.

DISCUSSION

In this paper, we used a SR-DBN and ELM for the classification of AD, MCI, and CN. In **Table 2** and **Figure 3**, the superiority of our model compared to other models can be seen, as evidenced by the highest values for accuracy, sensitivity, specificity, and AUC.

Table 3 presents a comparison of our model with previous deep learning models from the literature. Hu's model, Liu's model, and Suk's model adopted the stacked autoencoders and softmax classifier to classify AD. The thickness of the method

TABLE 2 | Results of the experiment.

		Proposed	Experiment 1	Experiment 2	Experiment 3	Experiment 4
AD vs. NC	ACC (%)	91.68 ± 1.09	85.68 ± 1.6	86.28 ± 1.02	76.43 ± 0.5	77.12 ± 0.04
	SEN (%)	95.47 ± 1.73	90.58 ± 2.43	90.92 ± 1.19	80.01 ± 0.62	74.02 ± 0.08
	SPE (%)	86.68 ± 2.61	79.6 ± 3.81	80.31 ± 2.24	71.99 ± 0.76	79.53 ± 0.13
	AUC	0.87 ± 0.01	0.82 ± 0.01	0.83 ± 0.02	0.75 ± 0.02	0.77 ± 0.03
MCIvsNC	ACC (%)	88.25 ± 1.38	80.34 ± 1.79	80.18 ± 1.03	67.64 ± 0.54	63.15 ± 0.02
	SEN (%)	79.74 ± 3.44	68.85 ± 4.77	68.33 ± 3.16	45.72 ± 1.1	78 ± 0.09
	SPE (%)	91.58 ± 2.14	87.19 ± 2.73	87.22 ± 1.55	80.05 ± 0.53	41.96 ± 0.11
	AUC	0.79 ± 0.01	0.72 ± 0.03	0.73 ± 0.03	0.60 ± 0.04	0.60 ± 0.02
ADvs MCI	ACC (%)	80.35 ± 1.8	73.07 ± 2.52	71.95 ± 1.61	65.77 ± 0.6	63.71 ± 0.03
	SEN (%)	85.65 ± 3.63	76.3 ± 3.93	78.51 ± 2.42	74.49 ± 0.72	79.44 ± 0.08
	SPE (%)	72.98 ± 4.71	69.41 ± 5.79	63.33 ± 4.06	54.15 ± 1.07	41.27 ± 0.09
	AUC	0.71 ± 0.08	0.69 ± 0.01	0.68 ± 0.02	0.59 ± 0.04	0.6 ± 0.02

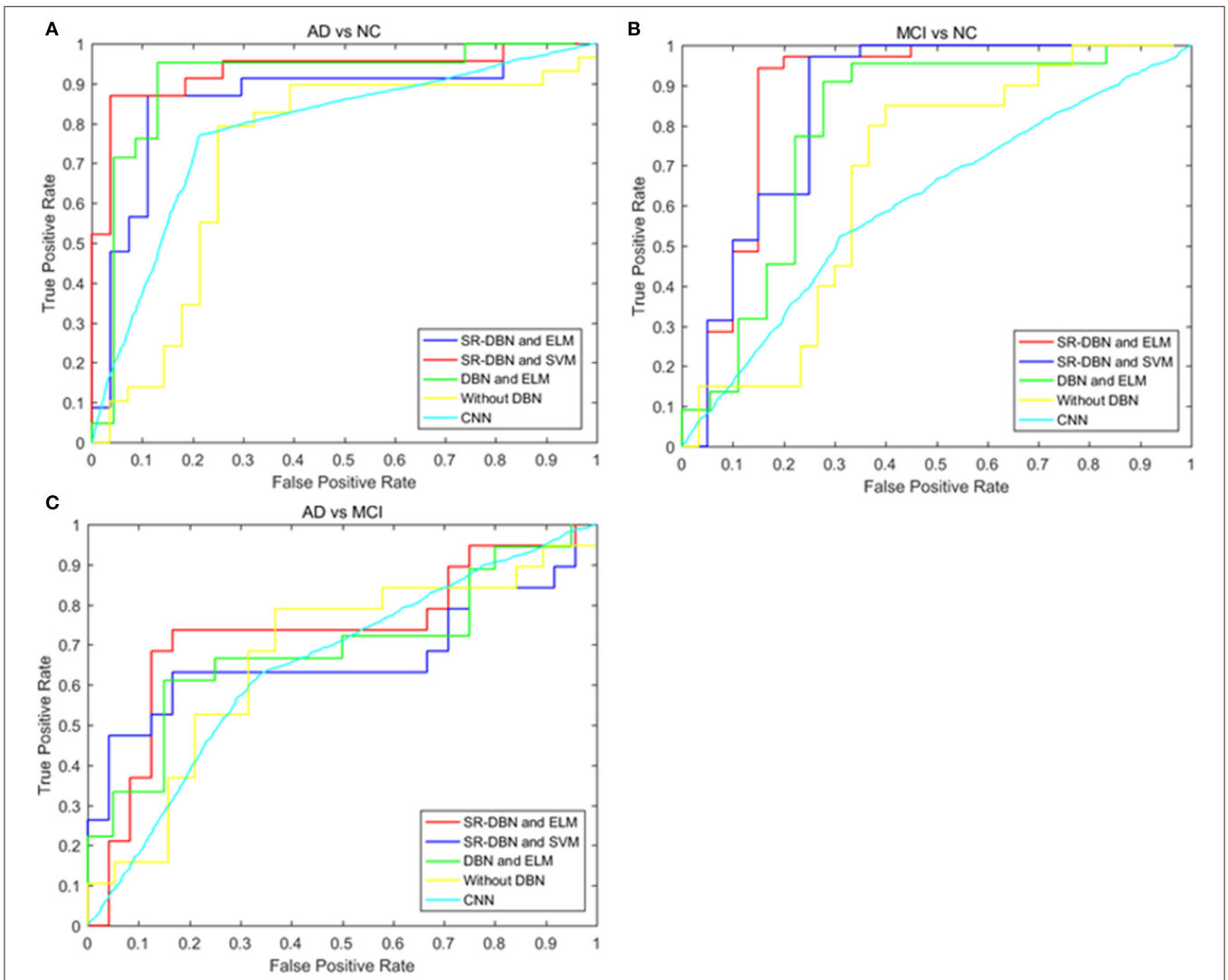


FIGURE 3 | The ROC curves of five models in the three experiments. **(A)** Shows the ROC curves of the classification of AD and NC; **(B)** shows the ROC curves of the classification of MCI and NC; **(C)** shows the ROC curves of the classification of AD and MCI.

TABLE 3 | Comparison of classification performances (%).

	AD vs. NC			MCI vs. NC		
	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)
Our model	91.68	95.47	86.68	88.25	79.74	91.58
Hu's model	-	-	-	87.5	-	-
Liu's model	87.76	88.57	87.22	76.92	74.29	78.13
Suk's model	83.2	-	-	70.1	-	-

likely contributed to gradient diffusion and gradient explosion, which was successfully avoided by using CD in our model. In addition, Hu's model used a single imaging modality (MRI) and Liu's model was a multimodal example. As shown in **Table 3**, the performance of our model was superior to the two models, reflecting the potential utility of our model to aid in early AD diagnosis.

However, the study has several limitations. Firstly, the parameters of the model ought to be modified to obtain better performance. Secondly, the method is based on multimodal data, but subjects with missing image data points are excluded, limiting the sample size. Thirdly, we only compared classification results among our proposed SR-DBN model, machine learning models, and classical CNN models in our dataset. Other deep learning models, such as recurrent neural networks, and deep neural network models were not compared in the same dataset. They will be implemented and compared in the future. Finally, the data used were from Western patients, which could potentially affect the results. Data from Eastern patients should be included in future studies to optimize the model and make it more generalizable to Eastern populations.

CONCLUSION

In the study, we proposed a SR-DBN combined with ELM to classify AD, MCI, and NC. Our model achieved 91.68% accuracy, 95.47% sensitivity, 86.68% specificity, and an AUC of 0.87 on the classification between AD and NC participants; 87.25% accuracy, 79.74% sensitivity, 91.58% specificity, and an AUC of 0.79 on classification between MCI and NC participants; and 80.35% accuracy, 85.65% sensitivity, 72.98% specificity, and an AUC of 0.71 on the classification between AD and MCI patients. Our model obtained better classification compared other models examined, indicating its effectiveness.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database.

AUTHOR CONTRIBUTIONS

PZ, SJ, and LY are responsible for writing experimental procedures, organizing experimental results and writing the paper. YF, CC, FL, and YL are responsible for experimental data collection and preprocessing. ZH was responsible for proposing experimental plans and guiding the writing of the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by grants; this program is sponsored by scientific development projects from ChenZhou Municipal Science and Technology Bureau (No. yfzx201906).

Data collection and dissemination for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI): the National Institutes of Health (grant number U01 AG024904), and the Department of Defense (award number W81XWH-12-2-0012). ADNI was funded by the National Institute of Aging and the National Institute of Biomedical Imaging and Bioengineering as well as through generous contributions from the following organizations: AbbVie, Alzheimer's Association, Alzheimer's Drug Discovery Foundation, Araclon Biotech, BioClinica Inc., Biogen, Bristol-Myers Squibb Company, CereSpir Inc., Eisai Inc., Elan Pharmaceuticals Inc., Eli Lilly and Company, EuroImmun, F. Hoffmann-La Roche Ltd. and its affiliated company Genentech Inc., Fujirebio, GE Healthcare, IXICO Ltd., Janssen Alzheimer Immunotherapy Research & Development LLC., Johnson & Johnson Pharmaceutical Research & Development LLC., Lumosity, Lundbeck, Merck & Co. Inc., Meso Scale Diagnostics LLC., NeuroRx Research, Neurotrack Technologies, Novartis Pharmaceuticals Corporation, Pfizer Inc., Piramal Imaging, Servier, Takeda Pharmaceutical Company, and Transition Therapeutics. The Canadian Institutes of Health Research are providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study was coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego, CA, USA. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California, CA, USA.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to EditSprings (https://www.editsprings.com/) for the expert linguistic services provided.

REFERENCES

1. Zhu X, Suk HI, Wang L, Lee SW, Initiative AsDN. A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. *Med Image Anal.* (2015) 75:570–7. doi: 10.1016/j.media.2015.10.008
2. Hua X, Leow AD, Parikshak N, Lee S, Chiang M-C, Toga AW, et al. Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease:

an MRI study of 676 AD, MCI, and normal subjects. *Neuroimage.* (2008) 43:458–69. doi: 10.1016/j.neuroimage.2008.07.013
3. Mosconi L, Tsui WH, Herholz K, Pupi A, Drzezga A, Lucignani G, et al. Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. *J Nucl Med.* (2008) 49:390–8. doi: 10.2967/jnumed.107.045385
4. Hu C, Ju R, Shen Y, Zhou P, Li Q. Clinical decision support for Alzheimer's disease based on deep learning and brain network. In: *2016 IEEE International Conference on Communications (ICC)*. (2016).
5. Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D. Early diagnosis of Alzheimer's disease with deep learning. In: *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. (2014).
6. Shen T, Jiang JH, Lin W, Ge JJ, Wu P, Zhou YJ, et al. Use of overlapping group LASSO sparse deep belief network to discriminate Parkinson's disease and normal control. *Front Neurosci.* (2019) 13:396. doi: 10.3389/fnins.2019.00396
7. Wang M, Jiang JH, Yan ZZ, Alberts I, Ge JJ, Zhang HW, et al. Individual brain metabolic connectome indicator based on Kullback-Leibler divergence similarity estimation predicts progression from mild cognitive impairment to Alzheimer's dementia. *Eur J Nucl Med Mol Imaging.* (2020) 47:2753–64. doi: 10.1007/s00259-020-04814-x
8. Ranzato MA, Boureau Y-L, Cun YL. Sparse feature learning for deep belief networks. *Adv Neural Inf Process Syst.* (2007) 20:1185–92.
9. Hinton GE. Training products of experts by minimizing contrastive divergence. *Neural Comput.* (2002) 14:1771–800. doi: 10.1162/089976602760128018
10. Ji N-N, Zhang J-S, Zhang C-X. A sparse-response deep belief network based on rate distortion theory. *Pattern Recognit.* (2014) 47:3179–91. doi: 10.1016/j.patcog.2014.03.025
11. Liang D, Pan P. Research on intrusion detection based on improved DBN-ELM. In: *2019 International Conference on Communications, Information System and Computer Engineering (CISCE)*, Haikou (2019).
12. Ribeiro B, Lopes N. Extreme learning classifier with deep concepts. In: *Iberoamerican Congress on Pattern Recognition*, Havana (2014).
13. Zhang D, Shen D, Initiative AsDN. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage.* (2012) 59:895–907. doi: 10.1016/j.neuroimage.2011.09.069
14. Gonzalez-Escamilla G, Lange C, Teipel S, Buchert R, Grothe MJ, Initiative A, et al. PETPVE12: an SPM toolbox for partial volume effects correction in brain PET–application to amyloid imaging with AV45-PET. *Neuroimage.* (2017) 147:669–77. doi: 10.1016/j.neuroimage.2016.12.077
15. Teipel SJ, Kurth J, Krause B, Grothe MJ, Initiative AsDN. The relative importance of imaging markers for the prediction of Alzheimer's disease dementia in mild cognitive impairment—beyond classical regression. *Neuroimage Clin.* (2015) 8:583–93. doi: 10.1016/j.nicl.2015.05.006
16. Suk HI, Shen D. Deep learning-based feature representation for ad/mci classification. *Med Image Comput Comput Assist Interv.* (2013) 16(Pt 2) 583–90. doi: 10.1007/978-3-642-40763-5_72

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 26 February 2021
doi: 10.3389/fmed.2021.628179

Evaluation of an Automatic Classification Algorithm Using Convolutional Neural Networks in Oncological Positron Emission Tomography

Pierre Pinochet[1], *Florian Eude*[1], *Stéphanie Becker*[1,2], *Vijay Shah*[3], *Ludovic Sibille*[3], *Mathieu Nessim Toledano*[1], *Romain Modzelewski*[1,2], *Pierre Vera*[1,2] *and Pierre Decazes*[1,2]*

[1] Department of Nuclear Medicine, Henri Becquerel Cancer Center, Rouen, France, [2] LITIS Quantif-EA 4108, University of Rouen, Rouen, France, [3] Siemens Medical Solutions USA, Inc., Knoxville, TN, United States

OPEN ACCESS

Edited by:
Xiaoli Lan,
Huazhong University of Science and Technology, China

Reviewed by:
Domenico Albano,
University of Brescia, Italy
Désirée Deandreis,
University of Turin, Italy

***Correspondence:**
Pierre Decazes
pierre.decazes@chb.unicancer.fr

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: 11 November 2020
Accepted: 25 January 2021
Published: 26 February 2021

Citation:
Pinochet P, Eude F, Becker S, Shah V, Sibille L, Toledano MN, Modzelewski R, Vera P and Decazes P (2021) Evaluation of an Automatic Classification Algorithm Using Convolutional Neural Networks in Oncological Positron Emission Tomography. Front. Med. 8:628179. doi: 10.3389/fmed.2021.628179

Introduction: Our aim was to evaluate the performance in clinical research and in clinical routine of a research prototype, called positron emission tomography (PET) Assisted Reporting System (PARS) (Siemens Healthineers) and based on a convolutional neural network (CNN), which is designed to detect suspected cancer sites in fluorine-18 fluorodeoxyglucose (^{18}F-FDG) PET/computed tomography (CT).

Method: We retrospectively studied two cohorts of patients. The first cohort consisted of research-based patients who underwent PET scans as part of the initial workup for diffuse large B-cell lymphoma (DLBCL). The second cohort consisted of patients who underwent PET scans as part of the evaluation of miscellaneous cancers in clinical routine. In both cohorts, we assessed the correlation between manually and automatically segmented total metabolic tumor volumes (TMTVs), and the overlap between both segmentations (Dice score). For the research cohort, we also compared the prognostic value for progression-free survival (PFS) and overall survival (OS) of manually and automatically obtained TMTVs.

Results: For the first cohort (research cohort), data from 119 patients were retrospectively analyzed. The median Dice score between automatic and manual segmentations was 0.65. The intraclass correlation coefficient between automatically and manually obtained TMTVs was 0.68. Both TMTV results were predictive of PFS (hazard ratio: 2.1 and 3.3 for automatically based and manually based TMTVs, respectively) and OS (hazard ratio: 2.4 and 3.1 for automatically based and manually based TMTVs, respectively). For the second cohort (routine cohort), data from 430 patients were retrospectively analyzed. The median Dice score between automatic and manual segmentations was 0.48. The intraclass correlation coefficient between automatically and manually obtained TMTVs was 0.61.

Conclusion: The TMTVs determined for the research cohort remain predictive of total and PFS for DLBCL. However, the segmentations and TMTVs determined automatically by the algorithm need to be verified and, sometimes, corrected to be similar to the manual segmentation.

Keywords: positron emission tomography, convolutional neural network, diffuse large B cell lymphoma (DLBCL), artificial intelligence-AI, fluorodeoxyglucose (^{18}F-FDG)

INTRODUCTION

Positron emission tomography (PET) with fluorine-18 (^{18}F) fluorodeoxyglucose (FDG) has an important contribution in the diagnosis and the management of oncological pathologies by highlighting regions with a high glucidic metabolism (1).

PET can establish an initial staging of tumor lesions (2), enable treatment optimization, and evaluate treatment effectiveness or possible relapse (3–8). It also provides prognostic parameters in certain types of cancer, in particular in onco-hematology, such as the Deauville score, which evaluates the therapeutic response and is used in clinical routine, or the total metabolic tumor volume (TMTV) (9).

TMTV represents, generally on FDG PET, the volume of the entire cancerous disease. It is obtained by segmenting each diagnosed lesion. TMTV has been shown to be an independent prognostic factor in lymphoma (10). Recently, Albano et al. have shown its predictive nature on progression-free survival (PFS) in elderly Hodgkin's lymphoma (11) and mantle cell lymphoma (12), but also on total and PFS in Burkitt lymphoma (13) and cerebral lymphoma (14). However, this parameter has some limitations. The first is that the measurement is time-consuming to make, explained by the fact that each lesion must be segmented individually, a task that cannot be performed manually in clinical practice. The second is the absence of a standard method for the segmentation of hypermetabolic lesions, which is responsible for some variability in the determination of TMTV. Thus, a fixed threshold of SUV_{max} (for example 41% for lymphomas) for each lesion is frequently used (15). However, this may not be appropriate for all pathological foci, particularly in the case of heterogeneous tumor fixation and adjacent physiological volume with high uptake (16).

A problem frequently encountered during the interpretation and segmentation of the images is differentiating between benign physiological (e.g., brain, heart, liver, kidney, and bladder) or inflammatory foci, and pathological foci suspicious for cancerous lesions. This is particularly true for malignant tumors with a low avidity for glucose, unusual location, or small size or in the presence of attenuation and/or motion artifacts (17). Moreover, inflammatory or infectious foci, or even foci with a high physiological consumption of glucose may have a sufficiently high FDG uptake to make it not possible to eliminate a cancerous origin (18, 19).

Intra- and interobserver interpretation of FDG PET/computed tomography (CT) findings has a high level of agreement in studies involving single site and experienced readers for lymphoma, lung, and head and neck cancers (20–22). Widespread adoption of TMTV would be facilitated by tools to assist image interpretation and standardize results. Automatic segmentation has also proven to be a prerequisite for certain studies, particularly in the field of radiomics.

In recent years, several automatic segmentation methods have been developed. They can be divided into two main groups. The first is based on an ROI placed manually by the physician within which a threshold relative to SUV_{max} is applied (23–25). The resulting segmentation depends on the defined ROI and is generally not optimal. A second approach, which is less time-consuming and observer-independent, uses supervised machine learning to analyze PET/CT images (26). A research software prototype called PET Assisted Reporting System (PARS), based on convolutional neural networks (CNNs), has recently been developed by Siemens Healthineers to classify hypermetabolic foci into benign and malignant and to provide parameters such as TMTV, total lesion glycolysis (TLG), and Deauville score (27). With this algorithm, PET volumes of interest are first segmented by using a fixed thresholding algorithm. Each volume of interest is then evaluated independently by using a combination of PET and CT multiplanar reconstructions, PET maximum intensity projections (MIPs), and atlas positions to predict the anatomic localization of FDG foci. These are input to a CNN that determines whether a focus is suspicious for malignancy. The training and validation sets were carried out on cohorts of patients with either lung cancer or lymphoma. A first, internal evaluation of this tool showed good accuracy of the automatic segmentation of FDG positive foci, and also good sensitivity and specificity of the classification in staging patients with lung cancer and lymphoma compared with manual segmentation (27).

The aim of this study was to verify the performance of PARS in order to determine its usefulness in research and clinical routine.

METHOD

Study Design

This retrospective monocentric study included patients treated at the Henri Becquerel Cancer Center, Rouen, France. Two patient cohorts were analyzed: a first clinical research cohort composed of patients with diffuse large B-cell lymphoma (DLBCL), as TMTV is a well-known prognostic factor for this disease (10), and a second clinical routine cohort composed of patients selected at random and followed up for miscellaneous cancers to evaluate if an automatic measurement of TMTV can be performed in routine. All patients were over 18 years of age. The baseline PET/CT was analyzed for the DLBCL clinical research cohort.

For the routine clinical cohort, including patients with suspected or confirmed cancer, a baseline or a follow-up PET/CT was analyzed. The study was approved by the institutional review board (no. 1901B). Patients were informed about the use of anonymized data for research and their right to oppose this use. Fully anonymized data were used, and explicit consent was waived.

Research Cohort

Concerning the research cohort, 119 patients followed up for DLBCL were included between November 2004 and September 2014, and their initial FDG PET/CT was analyzed.

PET/CT scans were acquired on a Biograph 16 (Siemens Healthineers, Knoxville, TN, USA). Patients fasted for at least 4 h and were injected with FDG at an activity of 3.5 MBq/kg of body weight. Images were acquired 60 min after injection at 2.5 min per bed position. The manual segmentation of lesions was performed using semiautomatic software (Planet Onco, version 2.0, DOSIsoft®, Cachan, France). A volume of interest was set around each lesion on the PET images. Then a fixed threshold value of 41% of SUV_{max} was applied to define the volume for each segmented lesion. The volumes of all suspicious lesions in a particular patient were added to compute the TMTV. The manual segmentation was performed by two nuclear physicians for each patient (MT and FE). One of the manual segmentations (MT), chosen arbitrarily, was used for the calculation of the Dice scores. The average of the two TMTVs was used for all other calculations.

Five-year follow-up, including PFS and overall survival (OS), was available for this cohort.

Routine Cohort

Concerning the routine cohort, 430 patients referred for cancer assessment underwent routine thoraco-abdomino-pelvic or whole body PET/CT (according to the indication), and with at least one tumoral uptake, were included between August 2018 and February 2020.

PET/CT scans were acquired on GE 710 (General Electric, Milwaukee, WI, USA) or Biograph Vision 600 (Siemens Healthineers, Knoxville, TN, USA). Patients fasted for at least 4 h and were injected with FDG at a dose of 3.0 MBq/kg of body weight. Images were acquired 60 min after injection at 2 min per bed position (GE 710) or by continuous bed motion (Biograph Vision).

The manual segmentation of lesions was performed using another semiautomatic software (PET VCAR, General Electric®) during routine clinical activity by two different nuclear medicine physicians (PD and PP). A volume of interest was set around each lesion on the PET images according to an adaptive thresholding (28), manually adapted if necessary according to medical advice. After the database was gathered, a second reading was done in order to check and confirm the suspicious character of the different segmented foci. These values were added to compute the TMTV.

Data of the two cohorts of patients are summarized in **Table 1**.

Convolutional Neural Network Use

PET/CT images were analyzed using a software prototype called PARS (Siemens Healthineers, Knoxville, TN, USA). A cylindrical reference region was automatically placed in the center of descending thoracic aorta to measure the mean blood pool uptake (SUV_{BP}). Regions on PET images with SUV_{peak} greater than $SUV_{BP} + 2\ std_{SUVBP}$ were identified and segmented using 42% of local SUV_{max}. Only segmentations with volumes over 2 ml (research cohort) or 1 ml (routine cohort) were selected to be processed by the CNN, which specifies location and physiological or suspicious character of the different foci.

Statistical Analysis

For both cohorts, agreement between automatic and manual segmentations was characterized using the Dice score. Differences between TMTVs from PARS and manual segmentation were determined using intraclass correlation coefficient (ICC), notably for subgroups of more than 30 patients. Comparisons were also made by way of Bland–Altman plots.

The prognostic value for PFS and OS for both automatic and manual TMTVs was analyzed in the research cohort. Hazard ratios were calculated on continuous data. Receiver operating characteristic (ROC) curves were used to determine TMTV cutoff thresholds by Youden's index. Survival functions were computed by Kaplan–Meier analyses and used to estimate survival time statistics for low and high TMTV groups with log-rank tests.

RESULTS

Research Cohort

Concerning the research cohort, 119 patients were included in the analysis. The median age was 65.8 years. Ninety-three patients had stage 3 or 4 DLBCL according to the Ann Arbor classification. Thirty received first-line treatment with R-ACVBP (doxorubicin, cyclophosphamide, vindesine, bleomycin, prednisone regimen) and 89 with R-CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisone regimen). The ICC between the two manual TMTVs was 0.86 ($p < 0.001$), confirming the reproducibility of the segmentations. The median Dice score across all patients between the set of PARS ROI's labeled as suspicious and the set of manual ROI's was 0.65. The average Dice score was 0.52. The median $TMTV_{PARS}$ was 194.79 ml, maximum 1,821 ml, and minimum 0 ml. The median $TMTV_{manual}$ was 313.34 ml, maximum 3,304 ml, and minimum 8 ml (**Table 1** and **Supplementary Figure 1**). The ICC between PARS and manual TMTVs was 0.68 (**Table 1**). Concerning the Bland–Altman plot, the deviation from the mean between $TMTV_{manual}$ and $TMTV_{PARS}$ was +204 ml with a confidence interval of −554 to +963 ml (see **Figure 1A**).

After a median follow-up of 5 years, 60 patients presented a recurrence of the disease and 54 deceased. The 5-year survival rates were 49.6% for PFS and 54.6% for OS.

The area under the ROC curve for predicting PFS was 0.62 for $TMTV_{PARS}$ and 0.71 for $TMTV_{manual}$ (**Figures 2A,B**). The optimal cutoffs for predicting PFS were 223.09 ml for $TMTV_{PARS}$ and 327.14 ml for $TMTV_{manual}$. The 5-year PFS

TABLE 1 | Summary results from two patient cohorts.

Cancer type	Number	Frequency (%)	Mean Dice score	Median Dice score	Minimal Dice score	Maximal Dice score	ICC (*p*-value)
Clinical research database							
Lymphoma (DLBCL)	119	100	0.52	0.65	0	0.93	0.68 ($p < 0.001$)
Clinical routine database							
Lung	150	34.88	0.51	0.59	0	0.94	0.71 ($p < 0.001$)
Lymphoma	71	16.51	0.40	0.45	0	0.97	0.56 ($p < 0.001$)
Breast	28	6.51	0.21	0	0	0.87	
Unknown	26	6.05	0.30	0.04	0	0.93	
Colorectal	25	5.81	0.46	0.56	0	0.88	
Melanoma	23	5.35	0.37	0.50	0	0.85	
Head and neck	23	5.35	0.40	0.36	0	0.93	
Esophagus	19	4.42	0.54	0.74	0	0.90	
Ovary	8	1.86	0.51	0.61	0	0.94	
Anal	5	1.16	0.10	0	0	0.46	
Kidney	5	1.16	0.24	0.22	0	0.61	
Sarcoma	5	1.16	0.22	0	0	0.63	
Carcinoma of unknown primary (CUP)	4	0.93	0.28	0.30	0	0.51	
Cervix	4	0.93	0.16	0	0	0.64	
Endometrium	4	0.93	0.18	0.02	0	0.66	
Pancreas	4	0.93	0.35	0.27	0	0.86	
Prostate	4	0.93	0.24	0.07	0	0.82	
Pleural	3	0.70	0.21	0	0	0.63	
Adrenal	3	0.70	0.18	0	0	0.54	
Testis	3	0.70	0.19	0	0	0.56	
Skin	2	0.47	0.40	0.40	0	0.79	
Stomach	2	0.47	0.42	0.42	0.03	0.81	
Myeloma	2	0.47	0.43	0.43	0	0.87	
Bladder	2	0.47	0.85	0.85	0.85	0.85	
Liver	1	0.23	0.53	0.53	0.53	0.53	
Leukemia	1	0.23	0.74	0.74	0.74	0.74	
Paraganglioma	1	0.23	0.70	0.70	0.70	0.70	
Thymus	1	0.23	0.94	0.94	0.94	0.94	
Thyroid	1	0.23	0.20	0.20	0.20	0.20	
Miscellaneous	430	100.00	0.42	0.48	0	0.97	0.61 ($p < 0.001$)

DLBCL, diffuse large B-cell lymphoma.

rates were 61.5 and 35.2% for the low- and high-$TMTV_{PARS}$ groups and 69.8% and 26.8% for the low- and high-$TMTV_{manual}$ groups, respectively (**Figures 3A,B**). The log-rank test indicated a significantly longer PFS time in the low-TMTV group for both TMTV estimation methods ($p = 0.0034$ for $TMTV_{PARS}$ and $p < 0.0001$ for $TMTV_{manual}$). Hazard ratios (high-TMTV group vs. low-TMTV group) were 2.1 (range 1.3–3.5) for $TMTV_{PARS}$ and 3.3 (range 2.0–5.6) for $TMTV_{manual}$.

For OS, the area under the ROC curve was 0.66 for $TMTV_{PARS}$ and 0.71 for $TMTV_{manual}$ (**Figures 2C,D**). The optimal cutoffs for predicting OS were 220.80 ml for $TMTV_{PARS}$ and 327.14 ml for $TMTV_{manual}$. The 5-year OS rates were 68.3 and 39.3% for the low- and high-$TMTV_{PARS}$ groups and 73.0% and 33.9% for the low- and high-$TMTV_{manual}$ groups, respectively (**Figures 3C,D**). The log-rank test indicated a significantly longer PFS time in the low-TMTV group for both TMTV estimation methods ($p = 0.0016$ for $TMTV_{PARS}$ and $p = 0.0001$ for $TMTV_{manual}$). Hazard ratios (high-TMTV group vs. low-TMTV group) were 2.4 (range 1.4–4.1) for $TMTV_{PARS}$ and 3.1 (range 1.8–5.3) for $TMTV_{manual}$.

Routine Cohort

Concerning the routine cohort, 430 patients were analyzed; 35% of them had lung cancer, 17% lymphoma, 7% breast cancer, 6% colorectal cancer, 5% melanoma, 5% head and neck cancer, 4% esophageal cancer, and 15% another cancer. In 6% of the cases,

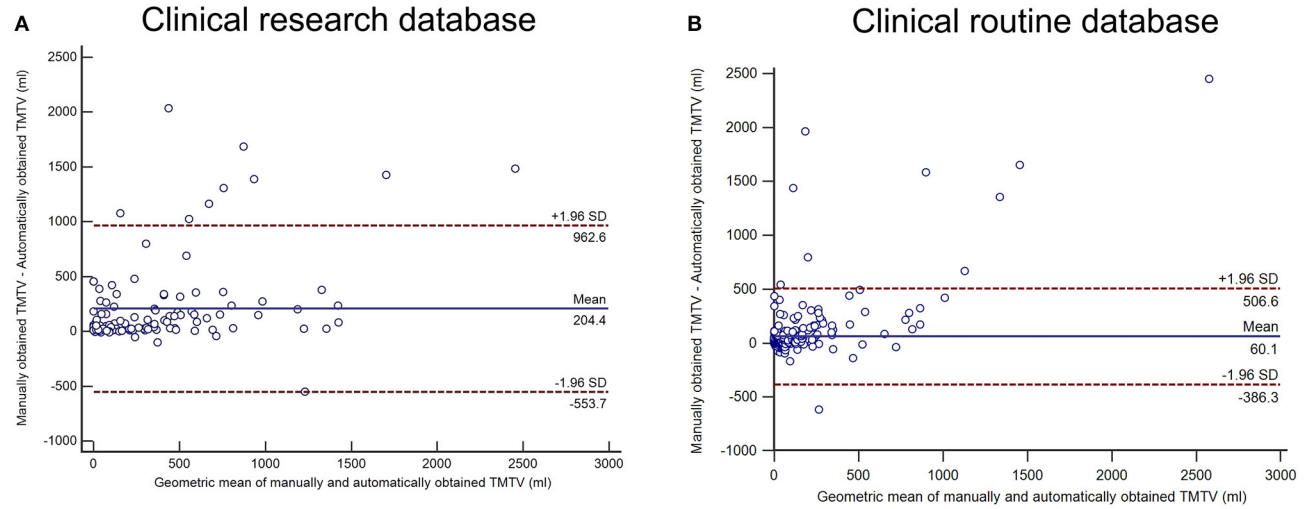


FIGURE 1 | Bland–Altman analysis between manually and automatically obtained total metabolic tumor volumes (TMTVs) for the clinical research cohort **(A)** and the clinical routine database **(B)**.

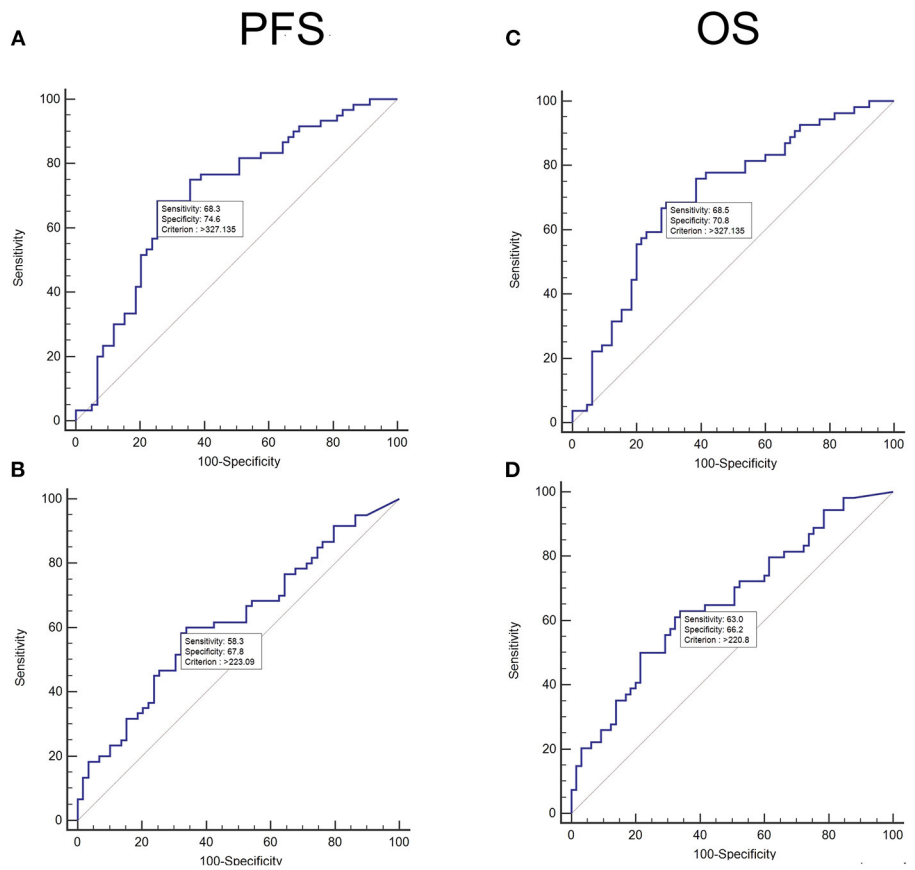


FIGURE 2 | Receiver operating characteristic (ROC) curve analysis of the population of diffuse large B-cell lymphomas (clinical research database) for progression-free survival (PFS) for manually obtained total metabolic tumor volumes (TMTVs) **(A)** and automatically obtained TMTVs **(B)** and for overall survival (OS) for manually obtained TMTVs **(C)** and automatically obtained TMTVs **(D)**.

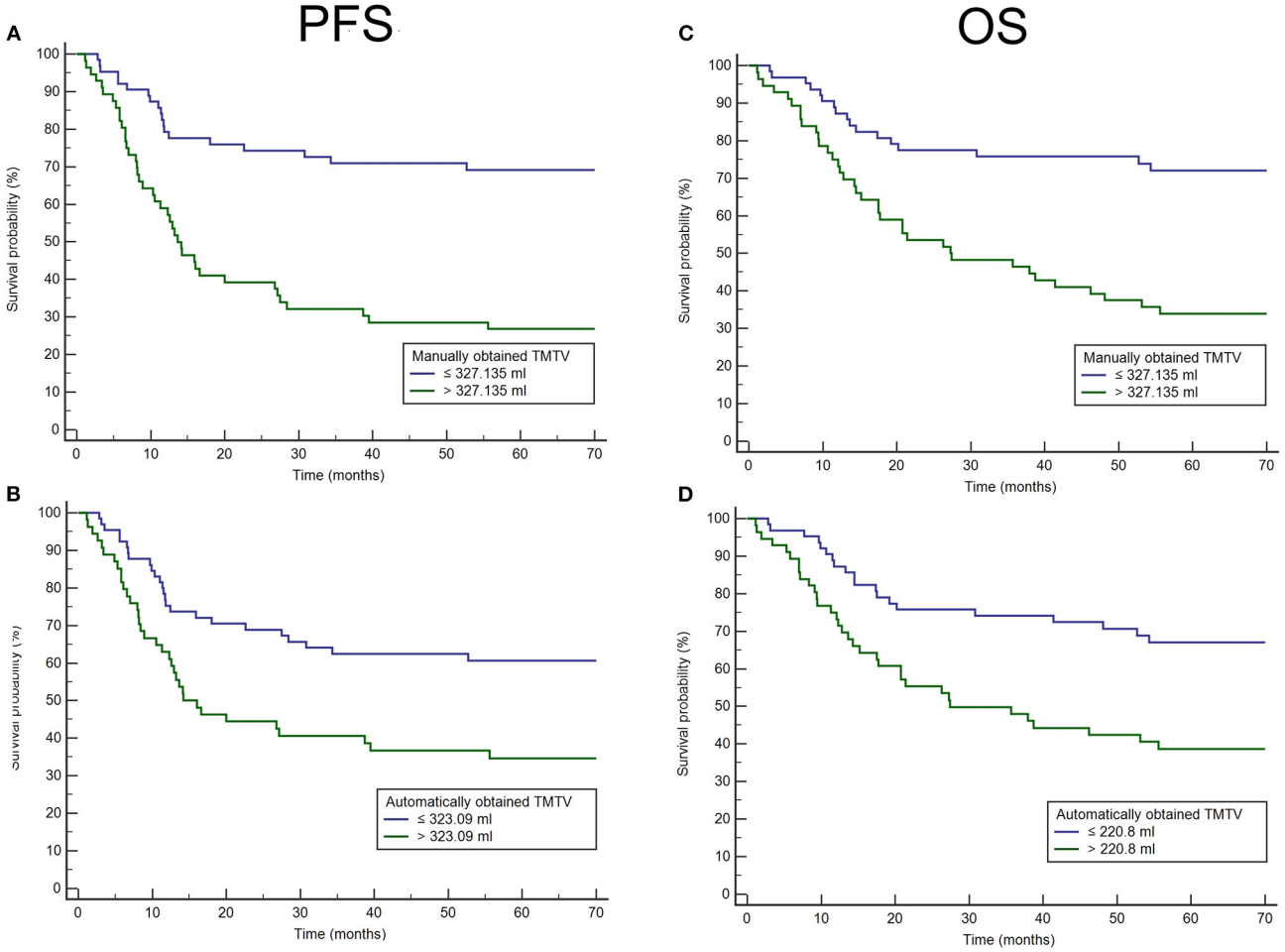


FIGURE 3 | Kaplan–Meier analysis of the population of diffuse large B-cell lymphomas (clinical research database) for progression-free survival (PFS) for manually obtained total metabolic tumor volumes (TMTVs) **(A)** and automatically obtained TMTVs **(B)** and for overall survival (OS) for manually obtained TMTVs **(C)** and automatically obtained TMTVs **(D)**.

the patients were followed up in another center, and we did not have the proven cancer origin.

The median Dice score across all patients between the suspicious PARS ROIs and the manual ROIs was 0.48. The average Dice score was 0.42. For automatic segmentation, median TMTV was 7.37 ml, maximum TMTV was 1,626.97 ml, and minimum TMTVs was 0.00 ml. For manual segmentation, median TMTV was 20.09 ml, maximum TMTV was 4,076.63 ml, and minimum TMTV was 1.00 ml (**Table 1** and **Supplementary Figure 1**). The intraclass coefficient between PARS and manual TMTV was 0.61 (**Table 1**). Concerning Bland–Altman plot, the deviation from the mean between $TMTV_{manual}$ and $TMTV_{PARS}$ was +60 ml with a confidence interval of −386 to +506 ml (see **Figure 1B**).

DISCUSSION

We analyzed an automatic segmentation software prototype using CNN in PET to distinguish hypermetabolic foci suspicious for cancer from nonsuspicious foci in two distinct cohorts of patients.

The first of these cohorts consisted of 119 patients with DLBCL, a disease used for the training of the model and for which the prognostic value of TMTV is well known (10). The median overlapping score of automatic and manual segmentation estimated by the Dice coefficient was 0.65. The ICC between automatically and manually determined TMTVs was 0.68. As follow-up was available for this cohort, survival analysis based on volume thresholds determined by the ROC curves showed that automatically determined TMTVs remained a predictive factor for PFS and OS, but hazard ratios were however lower than for manually determined TMTVs.

The second cohort consisted of 430 patients with a variety of cancers who were referred for PET/CT evaluation. The aim of the analysis of this cohort was to determine the possible utility of the algorithm for clinical routine, in terms of speed and reliability of the analysis of the different foci, and the estimation of the TMTVs. The median overlapping score of automatic and manual

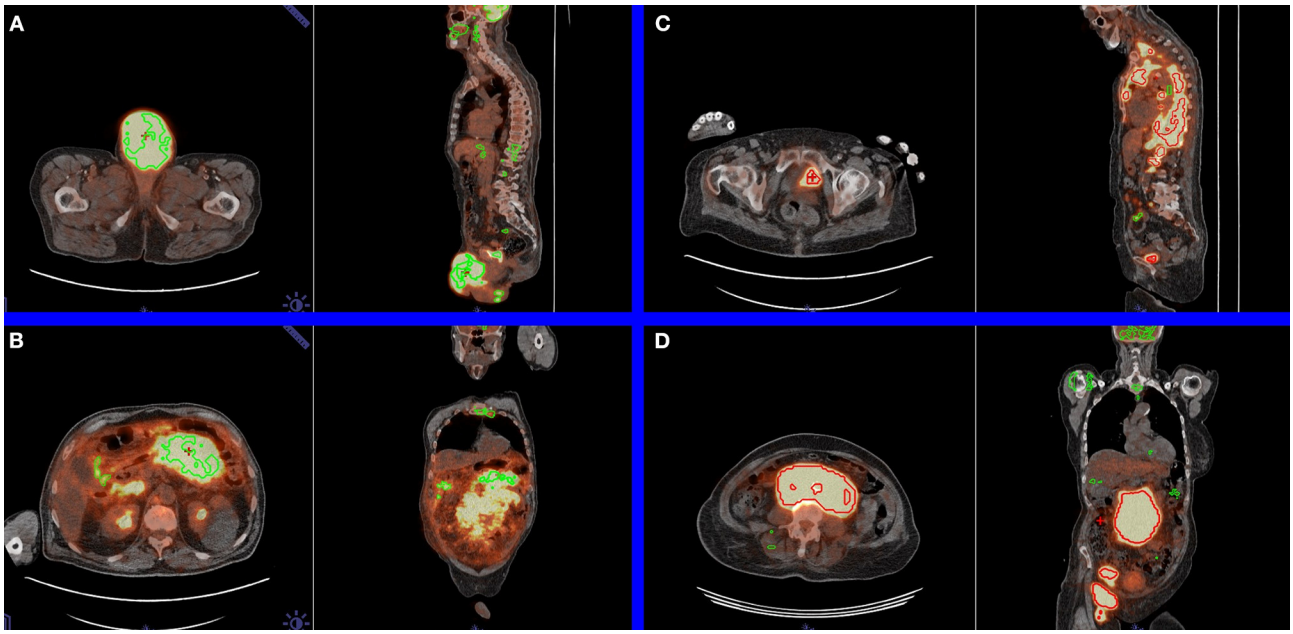


FIGURE 4 | Examples in axial and sagittal views of limitations of the automatic segmentation. **(A)** Pathological testicular mass labeled as physiological by positron emission tomography Assisted Reporting System (PARS) (false negative). For this patient, the manually and automatically obtained total metabolic tumor volumes (TMTVs) were 281.18 and 5.18 ml, respectively. **(B)** Pathological mesenteric mass was erroneously labeled as physiological by PARS (false negative). For this patient, the manually and automatically obtained TMTVs were 2,125.38 and 89.35 ml, respectively. **(C)** Physiological urinary bladder focus was erroneously labeled as pathological by PARS (false positive). For this patient, the manually and automatically obtained TMTVs were 816.94 and 661.08 ml, respectively. **(D)** Pathological mesenteric mass was correctly labeled as pathological by PARS (true positive). For this patient, the manually and automatically obtained TMTVs were 1,369.19 and 1,343.88 ml, respectively.

segmentation estimated by the Dice coefficient was 0.48. The ICC between automatically and manually determined TMTVs was 0.61.

The scanner type and acquisition parameters were different between the two cohorts. However, the results obtained were relatively similar despite these differences. Moreover, the manual segmentation methods differed (fixed threshold for the clinical research cohort and adaptive threshold for the routine cohort), but this did not greatly influence the results. The use of the 41% SUV_{max} thresholding method has been published in the context of DLBCLs and is a standard in clinical research (15), although much discussed (16). In particular, this method is difficult to use in clinical routine where tumor lesions are often smaller than those observed in DLBCL where a threshold of 41% of the SUV_{max} becomes unsuitable because of the partial volume effect for small lesions (29).

Finally, in the PARS configuration, to limit the computation time without impacting the TMTV measurement, only segmentations with volumes over 1 ml in the routine cohort were analyzed, as potentially small tumors were observed while the limit of 2 ml was used in the research cohort, as DLBCLs present generally large tumors.

In recent years, a number of algorithms have been developed that focus on PET segmentation, mainly in lymphoma, using different branches of artificial intelligence (30–32). In particular, machine learning using CNNs is a major advance in medical imaging. In PET, this technology stands to assist the nuclear physician's interpretation by facilitating, or even refining, the analysis. Concerning lymphomas, and DLBCL particularly, TMTV is usually not calculated during pretherapeutic PET/CT because it takes too long to determine using manual segmentation. Automatic or semiautomatic determination of TMTV could enable clinicians to integrate it in the determination of prognosis and therapeutic adaptation.

PARS is among the first published and validated CNN algorithms for PET/CT lesion classification (21). It was developed to detect FDG foci, and to predict the anatomic location and the expert classification (i.e., suspicious or not suspicious for cancer). It was trained on 380 examinations of patients with lung cancer or lymphoma with a validation set of 126 examinations and a test set of 123 patients (21).

In a recent study (33), the PARS software prototype was tested on a cohort of 280 patients with DLBCL. As with this study, we have established the ability to determine the prognosis of DLBCL using automatic segmentation. The authors however obtained a better lymphomatous lesion recovery coefficient (Dice) of 0.73 and a better TMTV correlation of 0.76. The automatically determined TMTVs were, as in our study, predictive of total and PFS with hazard ratios of 2.8 and 2.4, respectively. The difference in Dice coefficients and TMTV correlation could be explained by the difference in the populations.

Our results are consistent with a recent study (34), in which the performances of a CNN model, based on nnU-Net, were investigated to automatically segment TMTV in patients with DLBCL. A first cohort of 639 patients with pretherapeutic FDG PET/CT was used to train the model. In this cohort, the mean Dice score and Jaccard coefficients for manual and automatic segmentations were 0.73 and 0.68, respectively. There was a mean underestimation of automatic TMTV by 12 ml ($p = 0.27$). An external validation was done on a second cohort of 94 patients. In this testing set, the mean underestimation of automatically determined TMTV was 116 ml, which was statistically significant ($p = 0.01$).

Concerning the clinical routine database, we chose to analyze the examinations of patients followed for any cancerous pathology, whereas the model was trained only on lung cancer and lymphomas. This approach corresponds well to the clinical routine where the pathology is variable, and the results remain consistent with those of the research cohort. Nevertheless, the results are more similar to those obtained for the research cohort, which is closer to the training conditions of the algorithm.

Although promising, the PARS software prototype tends, in this study, to underestimate the number of cancerous foci, leading to some false-negative cases (see **Figure 4**). For both clinical research and clinical routine cohorts, the results obtained suggest that a manual check is still needed after the automatic segmentation.

CONCLUSION

The purpose of our study was to evaluate the software prototype PARS, which applies CNNs to detect carcinologically suspicious foci of hypermetabolism in FDG PET scans. The total tumor metabolic volumes determined by PARS were predictive of OS and PFS for patients belonging to the DLBCL research cohort. The segmentations and TMTVs determined automatically by the algorithm need to be verified and, sometimes, corrected to be similar to the manual segmentation in both clinical research and clinical routine.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Henri Becquerel Center Internal Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PD is the guarantor of the paper. PD, PV, RM, and PP designed the study. PD, SB, PP, and MT ensured inclusion and follow-up of patients. PP, FE, MT and PD managed imaging procedures. PP, PD and PV analyzed the data. LS and VS developed the software prototype. All authors contributed in drawing up the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.628179/full#supplementary-material

Supplementary Figure 1 | Boxplot showing the distribution of the Dice scores for the research cohort (DLBCL) and for the routine cohort (miscellaneous, lung cancer, and lymphoma).

REFERENCES

1. Kostakoglu L, Agress H, Goldsmith SJ. Clinical role of FDG PET in evaluation of cancer patients. *RadioGraphics.* (2003) 23:315–40. doi: 10.1148/rg.232025705
2. El-Galaly TC, Gormsen LC, Hutchings M. PET/CT for staging; past, present, and future. *Sem Nucl Med Medicine.* (2018) 48:4–16. doi: 10.1053/j.semnuclmed.2017.09.001
3. Oyen WJ, Bussink J, Verhagen AF, Corstens FH, Bootsma GP. Role of FDG-PET in the diagnosis and management of lung cancer. *Exp Rev Anticancer Ther.* (2004) 4:561–7. doi: 10.1586/14737140.4.4.561
4. Kandathil A, Iii RCS, Subramaniam RM. Lung cancer recurrence: ^{18} F-FDG PET/CT in clinical practice. *Am J Roentgenol.* (2019) 213:1136–44. doi: 10.2214/AJR.19.21227
5. Szyszko TA, Cook GJR. PET/CT and PET/MRI in head and neck malignancy. *Clin Radiol.* (2018) 73:60–9. doi: 10.1016/j.crad.2017.09.001
6. Kemppainen J, Hynninen J, Virtanen J, Seppänen M. PET/CT for evaluation of ovarian cancer. *Sem Nucl Med Medicine.* (2019) 49:484–92. doi: 10.1053/j.semnuclmed.2019.06.010
7. Gandy N, Arshad MA, Park W-HE, Rockall AG, Barwick TD. FDG-PET imaging in cervical cancer. *Sem Nucl Med Medicine.* (2019) 49:461–70. doi: 10.1053/j.semnuclmed.2019.06.007
8. Ulaner GA. PET/CT for patients with breast cancer: where is the clinical impact? *Am J Roentgenol.* (2019) 213:254–65. doi: 10.2214/AJR.19.21177
9. Cheson BD. PET/CT in lymphoma: current overview and future directions. *Sem Nucl Med.* (2018) 48:76–81. doi: 10.1053/j.semnuclmed.2017.09.007
10. Guo B, Tan X, Ke Q, Cen H. Prognostic value of baseline metabolic tumor volume and total lesion glycolysis in patients with lymphoma: a meta-analysis. *PLoS ONE.* (2019) 14:e0210224. doi: 10.1371/journal.pone.0210224
11. Albano D, Mazzoletti A, Spallino M, Muzi C, Zilioli VR, Pagani C, et al. Prognostic role of baseline 18F-FDG PET/CT metabolic parameters in elderly HL: a two-center experience in 123 patients. *Ann Hematol.* (2020) 99:1321–1330. doi: 10.1007/s00277-020-04039-w
12. Albano D, Bosio G, Bianchetti N, Pagani C, Re A, Tucci A, et al. Prognostic role of baseline 18F-FDG PET/CT metabolic parameters in mantle cell lymphoma. *Ann Nucl Med.* (2019) 33:449–58. doi: 10.1007/s12149-019-01354-9
13. Albano D, Bosio G, Pagani C, Re A, Tucci A, Giubbini R, et al. Prognostic role of baseline 18F-FDG PET/CT metabolic parameters in Burkitt lymphoma. *Eur J Nucl Med Mol Imaging.* (2019) 46:87–96. doi: 10.1007/s00259-018-4173-2
14. Albano D, Bertoli M, Battistotti M, Rodella C, Statuto M, Giubbini R, et al. Prognostic role of pretreatment 18F-FDG PET/CT in primary brain lymphoma. *Ann Nucl Med.* (2018) 32:532–41. doi: 10.1007/s12149-018-1274-8

15. Meignan M, Sasanelli M, Casasnovas RO, Luminari S, Fioroni F, Coriani C, et al. Metabolic tumour volumes measured at staging in lymphoma: methodological evaluation on phantom experiments and patients. *Eur J Nucl Med Mol Imaging.* (2014) 41:1113–22. doi: 10.1007/s00259-014-2705-y
16. Ilyas H, Mikhaeel NG, Dunn JT, Rahman F, Møller H, Smith D, et al. Defining the optimal method for measuring baseline metabolic tumour volume in diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imaging.* (2018) 45:1142–54. doi: 10.1007/s00259-018-3953-z
17. Griffeth LK. Use of Pet/Ct scanning in cancer patients: technical and practical considerations. *Baylor Univ Med Center Proc.* (2005) 18:321–30. doi: 10.1080/08998280.2005.11928089
18. Vaidyanathan S, Patel CN, Scarsbrook AF, Chowdhury FU. FDG PET/CT in infection and inflammation—current and emerging clinical applications. *Clin Radiol.* (2015) 70:787–800. doi: 10.1016/j.crad.2015.03.010
19. Rahman WT, Wale DJ, Viglianti BL, Townsend DM, Manganaro MS, Gross MD, et al. The impact of infection and inflammation in oncologic 18F-FDG PET/CT imaging. *Biomed Pharm.* (2019) 117:109168. doi: 10.1016/j.biopha.2019.109168
20. Hofman MS, Smeeton NC, Rankin SC, Nunan T, O'Doherty MJ. Observer variation in interpreting 18F-FDG PET/CT findings for lymphoma staging. *J Nuclear Med.* (2009) 50:1594–597. doi: 10.2967/jnumed.109.064121
21. Hofman MS, Smeeton NC, Rankin SC, Nunan T, O'Doherty MJ. Observer variation in FDG PET-CT for staging of non-small-cell lung carcinoma. *Eur J Nucl Med Mol Imaging.* (2009) 36:194–9. doi: 10.1007/s00259-008-0946-3
22. Senft A, de Bree R, Golding RP, Comans EFI, Van Waesberghe J-HTM, Kuik JD, et al. Interobserver variability in chest ct and whole body FDG-PET screening for distant metastases in head and neck cancer patients. *Mol Imaging Biol.* (2011) 13:385–90. doi: 10.1007/s11307-010-0354-5
23. Black QC, Grills IS, Kestin LL, Wong C-YO, Wong JW, Martinez AA, et al. Defining a radiotherapy target with positron emission tomography. *Int J Rad Oncol Biol Phys.* (2004) 60:1272–82. doi: 10.1016/j.ijrobp.2004.06.254
24. Tylski P, Stute S, Grotus N, Doyeux K, Hapdey S, Gardin I, et al. Comparative assessment of methods for estimating tumor volume and standardized uptake value in 18F-FDG PET. *J Nucl Med.* (2010) 51:268–76. doi: 10.2967/jnumed.109.066241
25. Vauclin S, Doyeux K, Hapdey S, Edet-Sanson A, Vera P, Gardin I. Development of a generic thresholding algorithm for the delineation of ^{18} FDG-PET-positive tissue: application to the comparison of three thresholding models. *Phys Med Biol.* (2009) 54:6901–16. doi: 10.1088/0031-9155/54/22/010
26. Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O. Deep learning with convolutional neural network in radiology. *Jpn J Radiol.* (2018) 36:257–72. doi: 10.1007/s11604-018-0726-3
27. Sibille L, Seifert R, Avramovic N, Vehren T, Spottiswoode B, Zuehlsdorff S, Schäfers M. ^{18} F-FDG PET/CT uptake classification in lymphoma and lung cancer by using deep convolutional neural networks. *Radiology.* (2020) 294:445–52. doi: 10.1148/radiol.2019191114
28. Doyeux K, Vauclin S, Hapdey S, Daouk J, Edet-Sanson A, Vera P, et al. Reproducibility of the adaptive thresholding calibration procedure for the delineation of 18F-FDG-PET-positive lesions. *Nucl Med Commun.* (2013) 34:432–8. doi: 10.1097/MNM.0b013e32835fe1f4
29. Berthon B, Marshall C, Edwards A, Evans M, Spezi E. Influence of cold walls on PET image quantification and volume segmentation: A phantom study: Influence of cold walls on PET image quantification and volume segmentation. *Med Phys.* (2013) 40:082505. doi: 10.1118/1.4813302
30. Yu Y, Decazes P, Gardin I, Vera P, Ruan S. 3D lymphoma segmentation in PET/CT images based on fully connected CRFs. In: Cardoso MJ, Arbel T, Gao F, Kainz B, van Walsum T, Shi K, Bhatia KK, Peter R, Vercauteren T, Reyes M, editors. *Molecular Imaging, Reconstruction Analysis of Moving Body Organs*. Cham: Springer International Publishing (2017). p. 3–12. doi: 10.1007/978-3-319-67564-0_1
31. Hu H, Decazes P, Vera P, Li H, Ruan S. Detection and segmentation of lymphomas in 3D PET images via clustering with entropy-based optimization strategy. *Int J CARS.* (2019) 14:1715–24. doi: 10.1007/s11548-019-02049-2
32. Yu Y, Decazes P, Lapuyade-Lahorgue J, Gardin I, Vera P, Ruan S. Semi-automatic lymphoma detection and segmentation using fully conditional random fields. *Comp Med Imag Grap.* (2018) 70:1–7. doi: 10.1016/j.compmedimag.2018.09.001
33. Capobianco N, Meignan MA, Cottereau A-S, Vercellino L, Sibille L, Spottiswoode B, et al. Deep learning FDG uptake classification enables total metabolic tumor volume estimation in diffuse large B-cell lymphoma. *J Nucl Med.* (2021). 62:30–6. doi: 10.2967/jnumed.120.242412
34. Blanc-Durand P, Jégou S, Kanoun S, Berriolo-Riedinger A, Bodet-Milin C, Kraeber-Bodéré F, et al. Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network. *Eur J Nucl Med Mol Imaging.* (2020). doi: 10.1007/s00259-020-05080-7. [Epub ahead of print].

Conflict of Interest: LS and VS are employees of the company Siemens Healthineers.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 28 April 2021
doi: 10.3389/fmed.2021.647562

A Preliminary Study to Use SUVmax of FDG PET-CT as an Identifier of Lesion for Artificial Intelligence

Kenji Hirata[1,2], *Osamu Manabe*[3,4*], *Keiichi Magota*[5], *Sho Furuya*[1,2], *Tohru Shiga*[1,2] *and Kohsuke Kudo*[1,6]

[1] Department of Diagnostic Imaging, Hokkaido University Graduate School of Medicine, Sapporo, Japan, [2] Department of Nuclear Medicine, Hokkaido University Hospital, Sapporo, Japan, [3] Department of Diagnostic and Interventional Radiology, Hokkaido University Hospital, Sapporo, Japan, [4] Department of Radiology, Saitama Medical Center, Jichi Medical University, Saitama, Japan, [5] Division of Medical Imaging and Technology, Hokkaido University Hospital, Sapporo, Japan, [6] Global Center for Biomedical Science and Engineering, Faculty of Medicine, Hokkaido University, Sapporo, Japan

Background: Diagnostic reports contribute not only to the particular patient, but also to constructing massive training dataset in the era of artificial intelligence (AI). The maximum standardized uptake value (SUVmax) is often described in daily diagnostic reports of [^{18}F] fluorodeoxyglucose (FDG) positron emission tomography (PET) – computed tomography (CT). If SUVmax can be used as an identifier of lesion, that would greatly help AI interpret diagnostic reports. We aimed to clarify whether the lesion can be localized using SUVmax strings.

Methods: The institutional review board approved this retrospective study. We investigated a total of 112 lesions from 30 FDG PET-CT images acquired with 3 different scanners. SUVmax was calculated from DICOM files based on the latest Quantitative Imaging Biomarkers Alliance (QIBA) publication. The voxels showing the given SUVmax were exhaustively searched in the whole-body images and counted. SUVmax was provided with 5 different degrees of precision: integer (e.g., 3), 1st decimal places (DP) (3.1), 2nd DP (3.14), 3rd DP (3.142), and 4th DP (3.1416). For instance, when SUVmax = 3.14 was given, the voxels with $3.135 \leq \text{SUVmax} < 3.145$ were extracted. We also evaluated whether local maximum restriction could improve the identifying performance, where only the voxels showing the highest intensity within some neighborhood were considered. We defined that "identical detection" was achieved when only single voxel satisfied the criterion.

Results: A total of 112 lesions from 30 FDG PET-CT images were investigated. SUVmax ranged from 1.3 to 49.1 (median = 5.6). Generally, when larger and more precise SUVmax values were given, fewer voxels satisfied the criterion. The local maximum restriction was very effective. When SUVmax was determined to 4 decimal places (e.g., 3.1416) and the local maximum restriction was applied, identical detection was achieved in 33.3% (lesions with $\text{SUVmax} < 2$), 79.5% ($2 \leq \text{SUVmax} < 5$), and 97.8% ($5 \leq \text{SUVmax}$) of lesions.

OPEN ACCESS

Edited by:
Kuangyu Shi,
University of Bern, Switzerland

Reviewed by:
Zhen Liu,
Huazhong University of Science and Technology, China
Jiaying Lu,
Fudan University, China

****Correspondence:***
Osamu Manabe
osamumanabe817@med.hokudai.ac.jp

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: *30 December 2020*
Accepted: *29 March 2021*
Published: *28 April 2021*

Citation:
Hirata K, Manabe O, Magota K, Furuya S, Shiga T and Kudo K (2021) A Preliminary Study to Use SUVmax of FDG PET-CT as an Identifier of Lesion for Artificial Intelligence. Front. Med. 8:647562. doi: 10.3389/fmed.2021.647562

Conclusion: In this preliminary study, SUVmax of FDG PET-CT could be used as an identifier to localize the lesion if precise SUVmax is provided and local maximum restriction was applied, although the lesions showing SUVmax < 2 were difficult to identify. The proposed method may have potential to make use of diagnostic reports retrospectively for constructing training datasets for AI.

Keywords: maximum of standardized uptake value, SUVmax, identifier, FDG PET, diagnostic report, artificial intelligence

INTRODUCTION

The clinical usefulness of positron emission tomography (PET) using [^{18}F]-fluorodeoxyglucose (FDG) has been well-established in oncology (1, 2). In addition to visual assessment (qualitative analysis), several quantitative measurements have been used to express the degree of FDG uptake. Among them, the standardized uptake value (SUV) has long been used as the de facto standard. To our knowledge, SUV was first extensively used around 1991 (3). In the initial years of its use, SUV was also known as the differential uptake ratio (4) or dose uptake ratio (5). The in-lesion maximum of SUV, or SUVmax, has frequently been used since 1999. By 2009 SUVmax had become the most frequently used measurement by far, with 6-fold more frequent use compared to the next most-often used measurement, according to a comprehensive review (6). Although SUV is most commonly calculated as the radioactivity concentration normalized to injection dosage and body weight, other definitions include the radioactivity concentration normalized to the body surface area (7), to lean body mass (6), and to blood glucose (8). In addition to SUVmax, metabolic tumor volume and total lesion glycolysis have been extensively investigated in recent studies and the evidence has been increasingly accumulated (9, 10). These volumetric measurements are, however, affected by the method of tumor boundary delineation, degrading inter-operator reproducibility. In contrast, SUVmax has theoretically highest inter-operator reproducibility. Many lines of evidence have demonstrated the usefulness of SUVmax for differential diagnosis, treatment response prediction, and prognosis (11).

Abbreviations: DP, decimal places; FDG, fluorodeoxyglucose; IQR, interquartile range; OSEM, ordered subsets expectation maximization; QIBA, Quantitative Imaging Biomarkers Alliance; SUV, standardized uptake value; VOI, volume of interest.

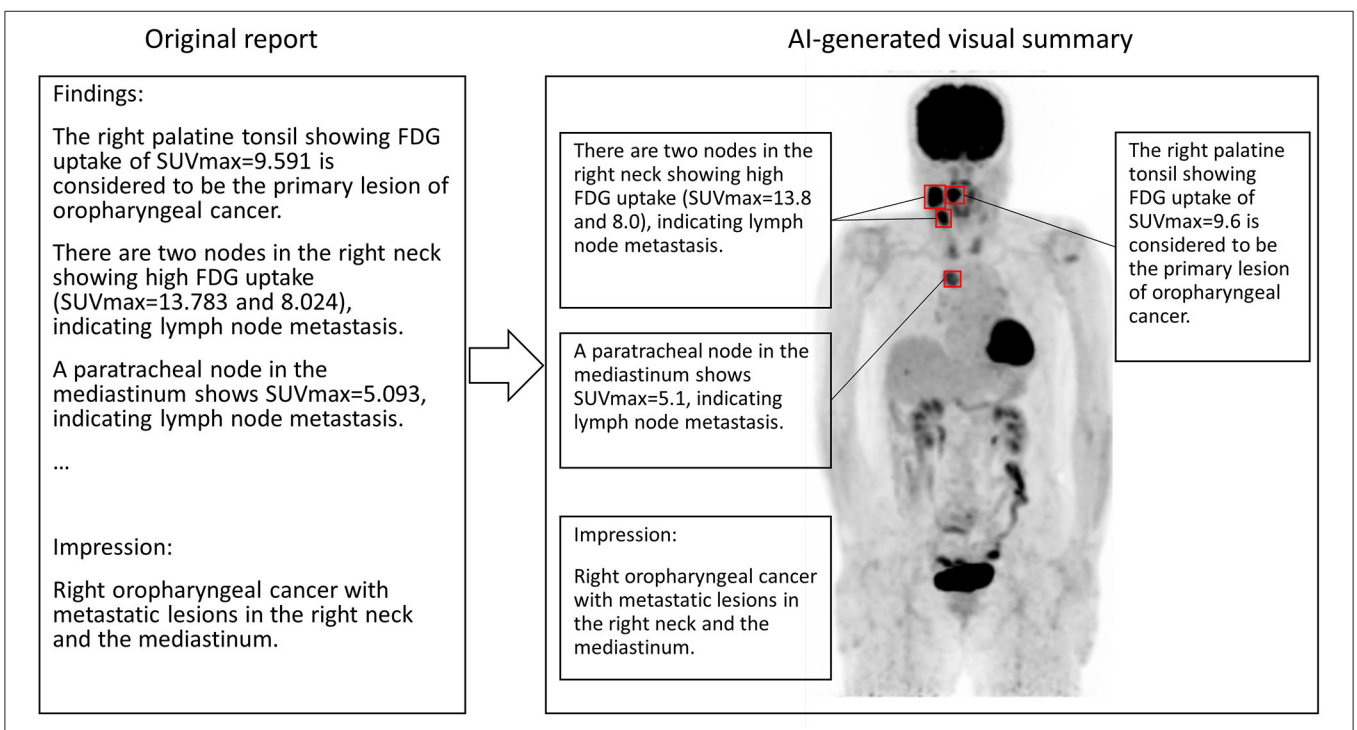


FIGURE 1 | A conceptual image of AI generating the visual summary of the report of FDG PET. SUVmax in the sentence appearing in the report text is used for localization. In this case, the primary lesion (right palatine tonsil) and metastatic nodes show high FDG uptakes. Note that SUVmax should be round before attending physicians read the report.

A total of 0.6 million FDG PET studies were performed in Japan in 2017 (12), and we think that diagnostic reports were written for most of them. Describing intensity of FDG uptake either using SUVmax or qualitatively has been recommended (13). Diagnostic reports not only contribute greatly to helping the attending physician interpret the image and diagnose the disease, but also prevent important findings from being overlooked. More recently, in the era of artificial intelligence (AI), the importance of training data is increasing. Collectively, diagnostic reports form a highly useful and efficient training database (14–18).

We hypothesized that if the SUVmax described in diagnostic reports was sufficiently precise, it might contribute to localization of the lesion, because there should be a limited number of voxels showing the same SUVmax in the entire image. In other words, we thought that SUVmax could be used as an identifier of the lesion. Thus, in this study, we aimed to clarify whether it would be possible to identify the lesion location using the SUVmax under various conditions by varying the degree of SUVmax precision and applying local maximum restriction. Such a technique could also be used to realize an automated system to generate a visual summary of the diagnostic report (**Figure 1**).

MATERIALS AND METHODS

Study Subjects

This retrospective observation study was approved by the institutional review board (approval no. 017-0454). The requirement of written informed consent from each patient was waived because of the study's retrospective nature. We confirmed that all methods were carried out in line with the relevant guidelines and regulations. A total of 30 PET-CT scans (sequential examinations for each scanner) were investigated in this study. No more than one scan was included for each patient. All the images were acquired between April 2019 and November 2019. Images were evaluated visually, and included to the study population if there were any pathological FDG uptakes in visual analysis until the number of scans reached 10 for each scanner. When all the FDG accumulation masses were considered physiological, the case was excluded. Note that not only uptake due to pathological malignancy but also malignancy-suspected and inflammatory uptakes were included in the analysis. In cases more than 5 uptakes were found, a maximum of 5 uptakes that showed highest values were recorded for a patient, based on

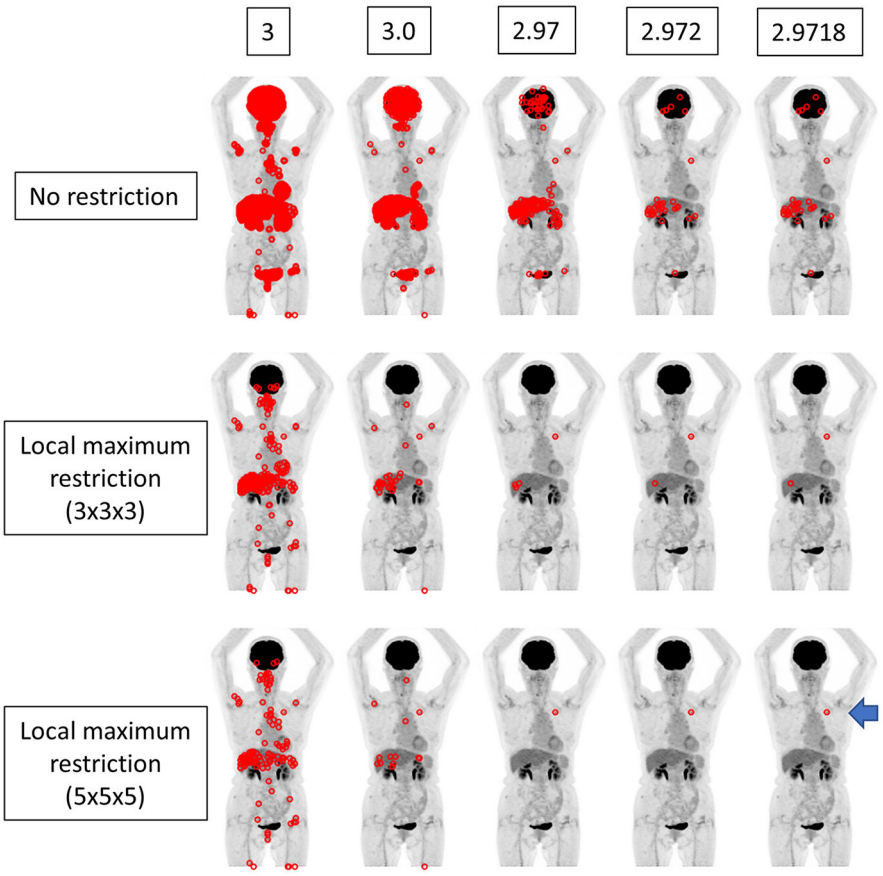


FIGURE 2 | The findings for a patient who underwent FDG PET-CT for lung nodules. The true SUVmax of the nodule in the left upper lobe was 2.97177 (arrow). When local maximum restriction was not applied, 21031, 2176, 210, 33, and 33 voxels were extracted for 3, 3.0, 2.97, 2.972, and 2.9718, respectively. When 3 × 3 × 3 local maximum restriction was applied, 254, 32, 4, 2, and 2 voxels were extracted. When 5 × 5 × 5 local maximum restriction was applied, 126, 14, 1, 1, and 1 voxel(s) were extracted, achieving identical detection.

RECIST 1.1 (19). Two experienced nuclear medicine physicians visually evaluated all the images. In case the interpretations of the two physicians differed, the final interpretation was reached by discussion.

PET-CT Image Acquisition and Reconstruction

In this study, we investigated images acquired with 3 different PET-CT scanners made by 2 different manufacturers.

Scanner 1 was a Biograph 64 True Point PET-CT (Siemens, Munich, Germany). The transaxial and axial fields of view were 68.4 and 21.6 cm, respectively. Emission data was acquired for 180 s per bed. Images were reconstructed using the OSEM algorithm with point spread function correction. Time-of-flight of photons was not measurable with the scanner. The reconstructed images had a matrix size of 168 × 168 and a voxel size of 4.1 × 4.1 × 2.0 mm.

Scanner 2 was a GEMINI TF64 PET-CT (Philips, Amsterdam, Netherlands). The transaxial and axial fields of view were 57.6 and 18.0 cm, respectively. Emission data was acquired for 60–180 s per bed depending on patient weight and injected dosage. Images were reconstructed using the OSEM algorithm reinforced with the time-of-flight algorithm. Point spread function correction was not applied. The reconstructed images had a matrix size of 144 × 144 and a voxel size of 4.0 × 4.0 × 4.0 mm.

Scanner 3 was a Vereos PET-CT (Philips, Amsterdam, Netherlands), which was the newest scanner of the three and equipped with digital photon counting detectors (20). The transaxial and axial fields of view were 67.6 and 16.4 cm, respectively (20). Emission data was acquired for 120–180 s per bed depending on patient weight and injected dosage. Images were reconstructed using the OSEM algorithm. Both the time-of-flight algorithm and point spread function correction were applied. The reconstructed images had a matrix size of 256 × 256 and a voxel size of 2.0 × 2.0 × 2.0 mm.

The number of voxels in the z-direction (i.e., cranio-caudal direction) ranged from 234 to 553, resulting in the final number of voxels ranging from 4.85×10^6 to 4.41×10^7. CT images were used for attenuation correction for all the scanners and

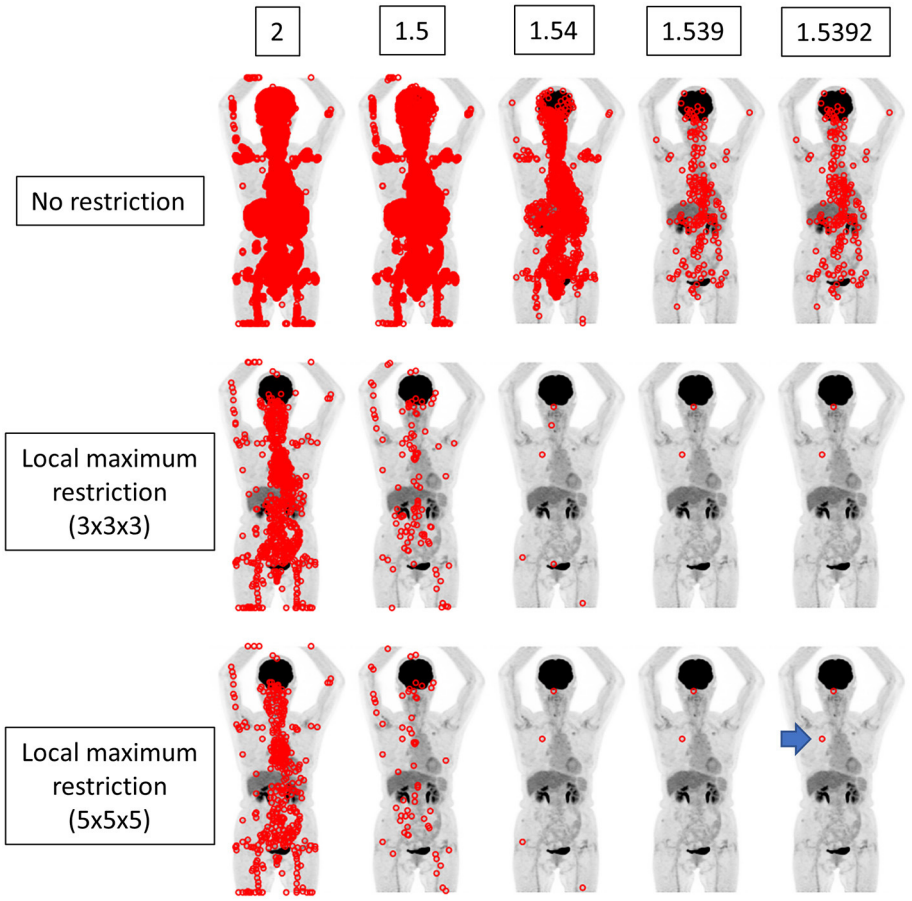


FIGURE 3 | The same case as depicted in **Figure 2**. The true SUVmax of the nodule in the right upper lobe was 1.53924 (arrow). When local maximum restriction was not applied, 74952, 13442, 1427, 198, and 198 voxels were extracted for 2, 1.5, 1.54, 1.539, and 1.5392, respectively. When 3 × 3 × 3 local maximum restriction was applied, 782, 104, 6, 2, and 2 voxels were extracted. When 5 × 5 × 5 local maximum restriction was applied, 410, 60, 4, 2, and 2 voxels were extracted. Thus, identical detection was not achieved for this lesion.

for visual assessment, but were not analyzed quantitatively in the current study. All patients fasted for ≥6 h before the injection of FDG (∼4 MBq/kg), and the emission scanning was initiated basically around 60 min post-injection. One scan was acquired 95 min post-injection due to mechanical troubles. Fasting blood sugar was confirmed to be smaller than 200 mg/dl in each study.

SUVmax Calculation

Commercially available DICOM viewers/PET viewers do not display SUVmax to 4 decimal places (DP) or higher. In order to obtain the ground truth of SUVmax, we modified Metavol software package, which we previously developed for PET-CT volumetric analysis (21). We used Windows 10, Microsoft Visual Studio Community 2019 Version 16.4.0, C# 8.0 language, .NET Core 3.1, and fo-dicom 4.0.3 for modifying Metavol. For instance, in the case that the true SUVmax is 3.14159, the modified Metavol will display it as it is, whereas XTREK VIEW software (J-MAC SYSTEM, Sapporo, Japan) will display it as 3.142. A nuclear medicine physician measured SUVmax by placing a spherical volume of interest (VOI) whose diameter can be changed by the operator. Another nuclear medicine physician independently confirmed all the values of SUVmax.

After the VOI definition, SUVmax was calculated based on the newest QIBA publication (22). Briefly, in Biograph64 and Vereos, the radioactivity concentration c (Bq/ml) was calculated as follows:

$$c = \rho \cdot s + i.$$

Here, ρ represents the raw pixel value that was stored with DICOM tag of (7FE0,0010) with each voxel expressed in a 16-bit integer. s represents the *rescale slope*, which is stored as a float value at (0028, 1053). i represents the *rescale intercept*, which is stored as a float value at (0028, 1052). Next, decay-corrected injection dosage D_c was calculated as follows:

$$D_c = D_0 \cdot (1/2)^{(\mathrm{Ta}-\mathrm{Ti})/\mathrm{h}}.$$

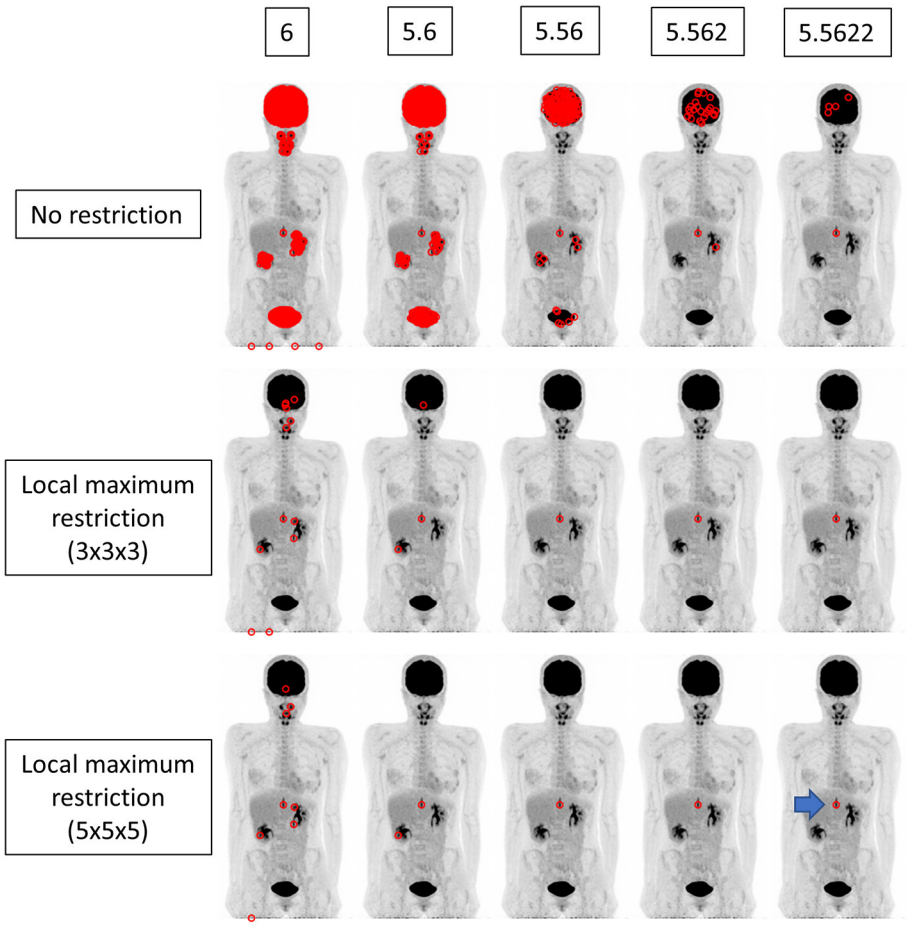


FIGURE 4 | The findings for a patient who underwent FDG PET-CT for a spinal code lesion. The true SUVmax of the nodule in the spinal code lesion was 5.56218 (arrow). When local maximum restriction was not applied, 21116, 1953, 186, 25, and 5 voxels were extracted for 6, 5.6, 5.56, 5.562, and 5.5622, respectively. When 3 × 3 × 3 local maximum restriction was applied, 12, 3, 1, 1, and 1 voxel(s) were extracted. When 5 × 5 × 5 local maximum restriction was applied, 8, 2, 1, 1, and 1 voxel(s) were extracted, achieving identical detection.

Here, D_0 represents the *radionuclide total dose* (i.e., injected dosage of FDG) (Bq) stored as a float value at (0018, 1074). T_a represents *acquisition time* stored at (0008, 0032). T_i represents the *radiopharmaceutical start time* (i.e., injection time) stored at (0018, 1072). Both times are stored in a "hhmmss" form string, and thus conversion to second is needed. h represents the *radionuclide half-life* (second) stored as a float value at (0018, 1075).

Finally, SUV was calculated as follows:

$$SUV = c \cdot w/D_c.$$

Here, w represents the *patient's weight* (g), which is stored in kilograms at (0010, 1030) and thus must be multiplied by 1000.

The SUV calculation was much simpler in GEMINI TF64, as follows:

$$SUV = (\rho \cdot s + i) \cdot p.$$

Here, p represents the *Philips Factor* (float) stored as a float value at (7053, 1000). The value of i was 0 for all the GEMINI TF64 examinations investigated in the current study.

Lesion Localization

We implemented a function that searches voxels satisfying the given SUV range and illustrate the locations in the whole body image (**Figures 2**–**4**). The SUV range was determined as follows. When "3" was provided by the operator, the range was considered to be $2.5 \leq SUV < 3.5$. When "3.1" was provided, the voxels satisfying $3.05 \leq SUV < 3.15$ were picked out, and so forth. Thus, the more precise the provided value of SUVmax (i.e., more digits) was, the narrower the range of SUV applied to extract voxels was. We compared the results from integer precision to 4th DP precision. Note that we do not show the results of 5th DP precision because there were no cases in which 5th DP precision improved the identification rate compared to 4th DP precision.

We performed experiments in different settings. First, the voxels within the range were extracted simply. Then, local maximum restriction was added to discard the voxel that was adjacent to the higher-value voxel, because such a voxel cannot have SUVmax. For local maximum restriction, milder restriction and stricter restriction were tested. Milder restriction was a condition under which the voxel must be highest in the $3 \times 3 \times 3$ cube. Stricter restriction was a condition under which the voxel must be highest in the $5 \times 5 \times 5$ cube.

Here, we defined that "identical detection" was achieved when only 1 voxel satisfied the criterion.

Statistical Analysis

The relationship between SUVmax vs. the number of voxels detected (N) was estimated using Pearson's correlation coefficient of the log of SUVmax vs. the log of N. The effects of the precision of SUVmax, i.e., the number of digits after the decimal point, and local maximum restriction on the rate of identical detection were tested using a chi-square test. P-values < 0.05 were considered statistically significant.

RESULTS

Patient characteristics are summarized in **Table 1**. Diagnosis and lesion locations are summarized in **Table 2**. In this study

TABLE 1 | Patient characteristics.

	Minimum	25-percentile	50-percentile (median)	75-percentile	Maximum
Age (year)*	11	62.25	69	75	86
Body weight (kg)	35.6	50.75	54.5	65.7	78.5
Fasting blood sugar (mg/dl)**	82	92.25	100.5	107	182
Injected dosage (MBq/body)	140.1	226.6	242.4	287.4	348.0
Injected dosage (MBq/kg)	2.97	4.33	4.41	4.47	4.74
Fasting time (hour)	5.5	7.0	15.5	17.0	22.0
Uptake time (min)***	53	55	56	60.5	95

1 (3%) patient was younger than 20 years old.
4 (13%) patients were diagnosed as having diabetes.
***Time duration between FDG injection and image acquisition start.*

TABLE 2 | Diagnosis and lesion sites.

Diagnosis	Number of patients	Site	Number of lesions
Head and neck cancer	11	Mediastinal and hilar nodes	29
Lung cancer	5	Bone	20
Colorectal cancer	4	Neck and subclavian nodes	17
Malignant lymphoma	2	Lung	16
Primary unknown cancer	2	Abdominal nodes	6
Spinal cord tumor	2	Nasal cavity and pharynx	4
Myelitis	1	Intestine	4
Hepatobiliary cancer	1	Breast	3
Mediastinal tumor	1	Spinal cord	3
Sarcoidosis	1	Other soft tissues	3
		Axillary nodes	2
		Liver	1
		Inguinal nodes	1
		Adrenal gland	1
		Parotid gland	1
		Peripheral nerve	1
Total	30		112

population, head-and-neck cancer was the most common diagnosis, and the mediastinal and hilar lymph nodes were the most frequent locations. In the 112 lesions investigated, SUVmax ranged from 1.3 to 49.1, with median and interquartile range (IQR) values of 5.6 and 5.2, respectively. SUVmax was significantly higher for Vereos than for Biograph64 and TF64 ($P < 0.01$ and $P < 0.05$, respectively; **Supplementary Figure 1**), which could be due to variability of diseases and scanner characteristics. The numbers of lesions for Biograph64, GEMINI TF64, and Vereos were 37, 37, and 38, respectively.

First, local maximum restriction was not applied. A number of voxels were identified corresponding to the given SUVmax (**Figure 5**, top row). Generally, when a larger SUVmax was given, a smaller number of voxels was detected ($0.83 < |r| < 0.84$, $P < 10^{-28}$). When the SUVmax was given with 10-fold greater precision, an ∼0.1-fold number of voxels were extracted, as expected theoretically.

Next, local maximum restriction was applied. Both 3 × 3 × 3 and 5 × 5 × 5 local maximum restriction reduced the number of extracted voxels up to 1/1000 (**Figure 5**, middle and bottom rows). More specifically, the rate of identical detection increased when the given SUVmax was more precise and local maximum restriction was stricter (**Figure 6**). For instance, while identical detection was successful only in 2.7% of patients when integer precision and no restriction were used, the success rate was elevated to 86.6% when 4th DP precision and 5 × 5 × 5 local maximum restriction were used. The effects of 5 × 5 × 5 over 3 × 3 × 3 local maximum restriction were observed as shown in **Figure 6**, except for integer precision, although none of the differences between 5 × 5 × 5 vs. 3 × 3 × 3 local maximum restriction reached the level of statistical significance ($P > 0.05$).

For sub-analysis, all lesions were categorized as low (SUVmax < 2, $N = 6$), medium ($2 \leq$ SUVmax < 5, $N = 44$), or high ($5 \leq$ SUVmax, $N = 62$) uptake lesions. The rate of identical detection was low (33.3%) for the low uptake group even under the best conditions, although the medium (79.5%) and high (96.8%) uptake groups achieved high rates (**Figure 7**). To investigate the underlying mechanisms for this difference, we drew a histogram of SUV over the whole-body image of a patient (**Figure 8**). In this case, the frequency exponentially decreased when SUVmax increased, as 98.13% of voxels showed $0 \leq$ SUV < 1, 1.28% showed $1 \leq$ SUV < 2, 0.37% showed $2 \leq$ SUV < 5, and 0.21% showed $5 \leq$ SUV.

In addition, **Table 3** summarizes the statistical analysis to search variables affecting the rate of identical detection of the lesion. In this analysis, all the lesions were categorized into 2 groups using the median of the variable as the cut-off. As the results, young age ($p = 0.01$) and large injected dosage (MBq/kg, $p = 0.005$) were significant factors for high rate of identical detection of the lesion. Note that injected dosage per body (MBq/body) was not a significant factor. We found that there were no significant correlations between SUVmax and patient age ($r = -0.01$) or between SUVmax and injected dosage (MBq/kg, $r = 0.17$).

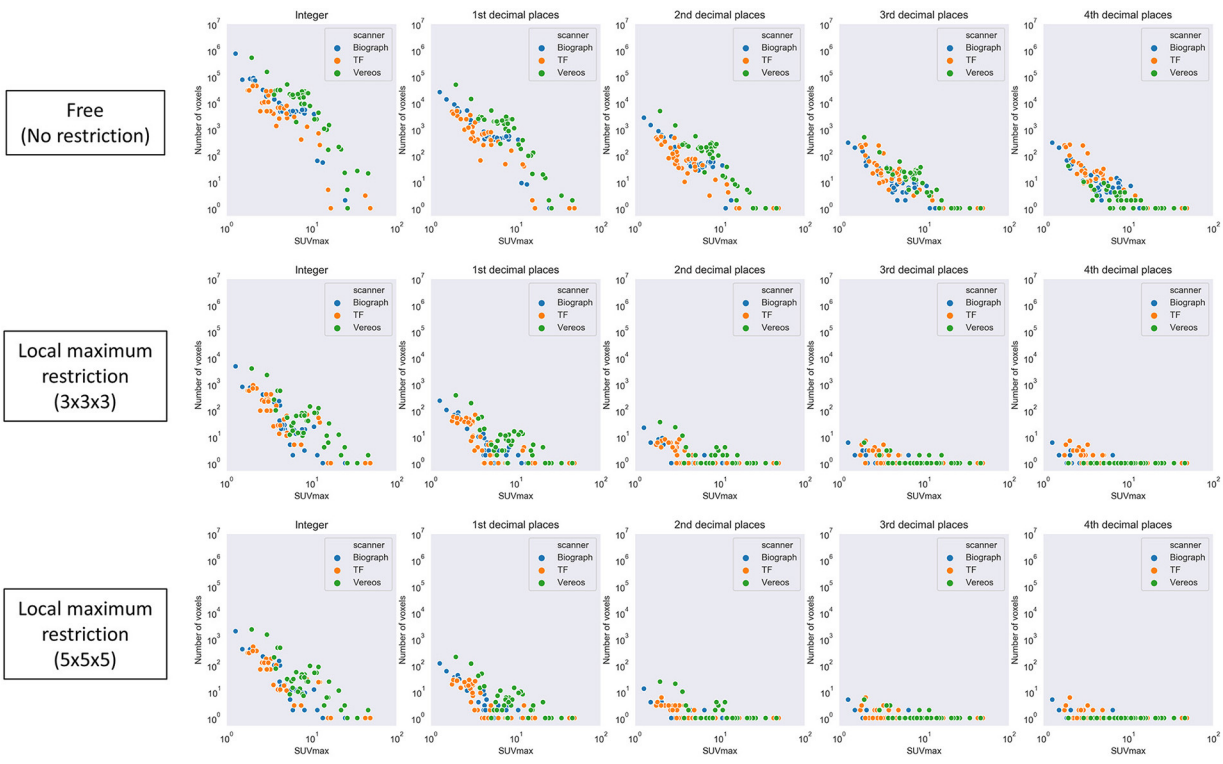


FIGURE 5 | The number of voxels extracted by a given SUVmax with various levels of precision. Top row, local maximum restriction was not applied; middle row, 3 × 3 × 3 local maximum restriction was applied; bottom row, 5 × 5 × 5 local maximum restriction was applied.

DISCUSSION

In this retrospective study, we aimed to clarify whether SUVmax can be used as a lesion identifier to localize the voxel in the whole-body image of FDG PET. We observed that SUVmax successfully localized the voxel for >80% examinations in the case that SUVmax was given to the 3rd or higher DP and local maximum restriction (5 × 5 × 5) was applied. However, the sub-analysis showed that the lesions having SUVmax < 2 were difficult to localize using SUVmax only. To our knowledge, this is the first report to show the use of SUVmax as an identifier of lesion on FDG PET-CT.

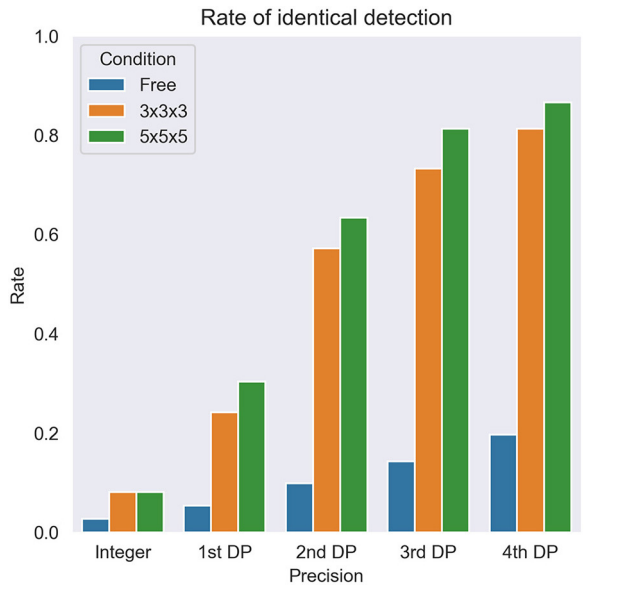


FIGURE 6 | The overall rate of identical detection of the lesion. DP, decimal places. Free, 3 × 3 × 3, and 5 × 5 × 5 express no restriction and each local maximum restriction.

The pixel data was stored in DICOM files in a 16-bit integer form for all 3 scanners investigated. A 16-bit integer can express 65,536 different values. Since the number of voxels in the whole-body image may be around 10^7, theoretically > 100 voxels on average may have exactly the same value. In fact, however, the distribution of SUV was quite skewed, as shown in **Figure 8**. It is reasonable that many voxels were detected when a smaller SUVmax was given, whereas only single voxel was detected when a larger SUV (e.g., >5) was given. In **Figure 5**, the number of voxels suddenly dropped once SUVmax became larger than 10. This can be explained as follows. In this study, we used DP instead of significant figures. They are slightly but clearly different. DP means the number of digits located to the right of the decimal point. Significant figures refers to the total number of digits irrespective of the decimal point location. For example, 9.8 is 1st DP and 2 significant figures, whereas 12.3 is 1st DP and 3 significant figures. Since 12.3 has more information than 9.8, fewer voxels were included within the range.

The effect of local maximum restriction was significant. The number of voxels that can become local maxima depends on the noise level of the image. Mathematically, when 3 × 3 × 3 restriction was applied, at most 1 of 2 voxels in each axis could become local maxima, indicating that 1/8 or a smaller number of voxels could become local maxima. Similarly, when 5 × 5 × 5 restriction was applied, at most 1 of 3 voxels in each axis and thus 1/27 or a smaller number of voxels could become local maxima. We did not try 7 × 7 × 7 restriction because we were worried that it might prevent identification of the voxel of SUVmax, considering that a single voxel size is 4 mm, and its diagonal is $4\sqrt{3} = 6.9$ mm, and thus 7 voxels account for as large as 48.5 mm.

Some may argue that use of the 3rd or higher DP for SUVmax is redundant for daily radiological reports. That is true. SUV calculation uses body weight and the precision of body weight measurement may be 3 significant figures (e.g., 56.7 kg) or less. Radioactivity dosage measurements may introduce some errors. Furthermore, SUV varies depending on various technical

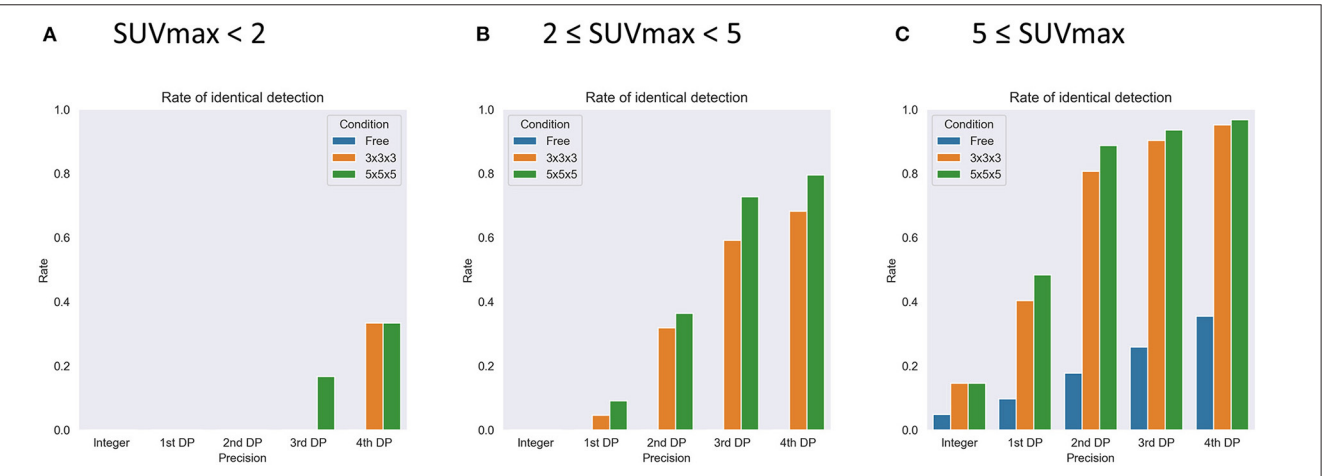


FIGURE 7 | The results of sub-group analysis of the rate of identical detection of the lesions with SUVmax < 2 (**A**, $N = 6$), 2 < SUVmax < 5 (**B**, $N = 44$), and 5 < SUVmax (**C**, $N = 62$). DP, decimal places. Free, 3 × 3 × 3, and 5 × 5 × 5 express no restriction and each local maximum restriction.

(e.g., scanner, acquisition protocol, reconstruction protocol) and physiological (e.g., fasting conditions) factors. Therefore, the number excessive fine number is meaningless in diagnosis and treatment planning. Those may be why SUVmax is often written to the 1st DP (e.g., 3.1). However, in order to permit the future use of SUVmax as an identifier, we would like to propose that SUVmax be written as precisely as the PET-CT viewer allows. As mentioned before, this use of SUVmax would allow the diagnostic report to be summarized as a single image (**Figure 1**). In addition, it may also help radiologists to locate a lesion mentioned in a previous report so as to compare between past and present images. Our ultimate goal is to build a massive training dataset based on diagnostic reports and corresponding images. Writing the coordinate values (x, y, z) in the reports will be the best way to transfer the information to artificial intelligence. Currently, that may not be possible in most viewers and reporting systems. Also, the appearance of such information in the middle of a report may distract readers, and thus an automated system is needed to hide this information when humans read the report.

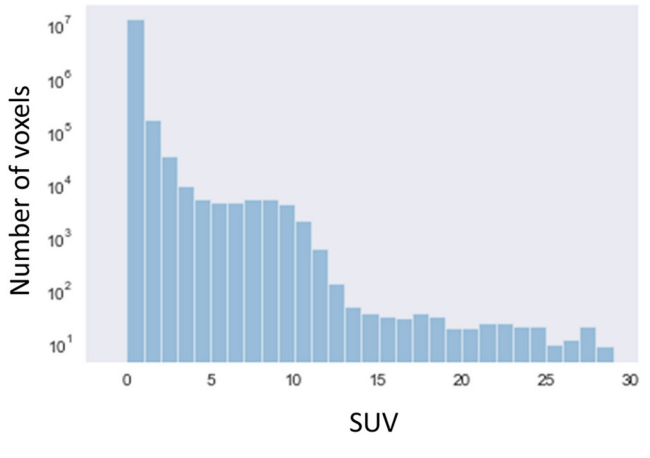


FIGURE 8 | A histogram of SUV over the whole-body image of a patient (semi-log plot).

In the sub-analysis, we observed that the success rate was affected by some other factors. More specifically, the success rate was higher when the patient was younger or the injected dosage (MBq/kg) was larger. To reveal the reasons, we calculated Pearson's correlation coefficients between SUVmax vs. these factors, but there were no significant correlations. Thus, the underlying mechanisms remain unclear and will be investigated in future studies.

Although some radiological reports may be written with SUVmean, we did not try using SUVmean as a lesion identifier. Technically, SUVmean can be used instead of SUVmax for our current method; however, SUVmean calculation for every location is a time-consuming process and thus may not be practical. In addition, the range of SUVmean is smaller than SUVmax, which makes SUVmean less feasible for a lesion identifier.

The use of SUVmax is specific for PET. Although the maximum voxel value may not often be useful for CT or MRI, the idea could be applied to the apparent diffusion coefficient (ADC) images derived from diffusion weighted imaging of MRI, because the minimum of ADC is meaningful for diagnosis.

As limitations of the current study, we did not investigate the SUVmax shown in different image viewers. In some viewers, PET volumes are reconstructed (resliced) in the CT alignment, making slight changes to SUVmax. Secondly, we did not directly use the diagnostic reports but reviewed the images to re-measure SUVmax. Thus, we could not estimate the number of actual cases in which the SUVmax written in the reports could successfully locate the lesion. Such a study needs to be carried out. Thirdly, we investigated only 30 cases for this preliminary study. A larger study will be needed to confirm the results. In addition, head and neck cancer accounted for a large portion of the current study population, which does not necessarily reflect general population undergoing PET-CT. Finally, diagnostic reports often provide anatomical terms in the same sentence with SUVmax. This would

TABLE 3 | Variables affecting the rate of identical detection of the lesion.

	<Median			≥Median			*p**
	Number of lesions (A)	Number of identical detection (B)	Rate (B/A)	Number of lesions (A)	Number of identical detection (B)	Rate (B/A)	
Age (year)	59	56	94.9%	53	41	77.4%	0.01
Body weight (kg)	58	52	89.7%	54	45	83.3%	0.41
Fasting blood sugar (mg/dl)	53	48	90.6%	59	49	83.1%	0.28
Injected dosage (MBq/body)	54	49	90.7%	58	48	82.8%	0.27
Injected dosage (MBq/kg)	51	39	76.5%	61	58	95.1%	0.005
Fasting time (hour)	47	41	87.2%	65	56	86.2%	1.0
Uptake time (min)**	36	30	83.3%	76	67	88.2%	0.56

p-values were calculated Fisher's exact test.

***Time duration between FDG injection and image acquisition start.*

be great information for selecting the appropriate location when SUVmax suggests several candidates, as in **Figure 3**. Such a method will be tested in future studies.

CONCLUSION

The data suggested that SUVmax can be used as an identifier of lesion on FDG PET-CT. For this purpose, it is important that SUVmax is given precisely (3rd DP or more) and that local maximum restriction is applied to identify the voxel. The lesions showing SUVmax < 2 were difficult to identify. As this is a preliminary study investigating a small population from a single center, a larger study with many more patients will be needed to validate the results. The proposed method may have potential to make use of diagnostic reports retrospectively for constructing training datasets for AI.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of Hokkaido University Hospital for Clinical Research. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements. All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional review board (017-0454) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards (EudraCT nr, 2017-003461-96).

AUTHOR CONTRIBUTIONS

KH and OM designed experiments, performed experiments, wrote and revised the manuscript. KM and SF collected image data. TS and KK wrote and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by JST COI Grant Number JPMJCE1301 (H31W09) and JSPS KAKENHI Grant Number JP20K08015.

ACKNOWLEDGMENTS

We thank all staff in the nuclear medicine department of our hospital.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.647562/full#supplementary-material

REFERENCES

1. Ben-Haim S, Ell P. 18F-FDG PET and PET/CT in the evaluation of cancer treatment response. *J Nucl Med.* (2009) 50:88–99. doi: 10.2967/jnumed.108.054205
2. Elsinga P, Delaloye AB, Chiti A, Vandenberghe S, Giammarile F, Carrio I. Endorsement of international consensus radiochemistry nomenclature guidelines. *Eur J Nucl Med Mol Imaging.* (2019) 46:1399. doi: 10.1007/s00259-018-4243-5
3. Haberkorn U, Strauss LG, Dimitrakopoulou A, Engenhart R, Oberdorfer F, Ostertag H, et al. PET studies of fluorodeoxyglucose metabolism in patients with recurrent colorectal tumors receiving radiotherapy. *J Nucl Med.* (1991) 32:1485–90.
4. Griffeth LK, Dehdashti F, McGuire AH, McGuire DJ, Perry DJ, Moerlein SM, et al. PET evaluation of soft-tissue masses with fluorine-18 fluoro-2-deoxy-D-glucose. *Radiology.* (1992) 182:185–94. doi: 10.1148/radiology.182.1.1727280
5. Adler LP, Blair HF, Makley JT, Williams RP, Joyce MJ, Leisure G, et al. Noninvasive grading of musculoskeletal tumors using PET. *J Nucl Med.* (1991) 32:1508–12.
6. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* (2009) 50 (Suppl. 1):122s−50s. doi: 10.2967/jnumed.108.057307
7. Kim CK, Gupta NC, Chandramouli B, Alavi A. Standardized uptake values of FDG: body surface area correction is preferable to body weight correction. *J Nucl Med.* (1994) 35:164–7.
8. Nozawa A, Rivandi AH, Kesari S, Hoh CK. Glucose corrected standardized uptake value (SUVgluc) in the evaluation of brain lesions with 18F-FDG PET. *Eur J Nucl Med Mol Imaging.* (2013) 40:997–1004. doi: 10.1007/s00259-013-2396-9
9. Kitao T, Shiga T, Hirata K, Sekizawa M, Takei T, Yamashiro K, et al. Volume-based parameters on FDG PET may predict the proliferative potential of soft-tissue sarcomas. *Ann Nucl Med.* (2019) 33:22–31. doi: 10.1007/s12149-018-1298-0
10. Kitao T, Hirata K, Shima K, Hayashi T, Sekizawa M, Takei T, et al. Reproducibility and uptake time dependency of volume-based parameters on FDG-PET for lung cancer. *BMC Cancer.* (2016) 16:576. doi: 10.1186/s12885-016-2624-3
11. Krak NC, Boellaard R, Hoekstra OS, Twisk JW, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *Eur J Nucl Med Mol Imaging.* (2005) 32:294–301. doi: 10.1007/s00259-005-1926-5
12. Nishiyama Y, Kinuya S, Kato T, Kayano D, Sato S, Tashiro M, et al. Nuclear medicine practice in Japan: a report of the eighth nationwide survey in 2017. *Ann Nucl Med.* (2019) 33:725–32. doi: 10.1007/s12149-019-01382-5
13. Niederkohr RD, Greenspan BS, Prior JO, Schoder H, Seltzer MA, Zukotynski KA, et al. Reporting guidance for oncologic 18F-FDG PET/CT imaging. *J Nucl Med.* (2013) 54:756–61. doi: 10.2967/jnumed.112.112177
14. Dalal S, Hombal V, Weng WH, Mankovich G, Mabotuwana T, Hall CS, et al. Determining follow-up imaging study using radiology reports. *J Digit Imaging.* (2019) 33:121–30. doi: 10.1007/s10278-019-00260-w
15. Liu Y, Liu Q, Han C, Zhang X, Wang X. The implementation of natural language processing to extract index lesions from breast magnetic resonance imaging reports. *BMC Med Inform Decision Making.* (2019) 19:288. doi: 10.1186/s12911-019-0997-3

16. Spandorfer A, Branch C, Sharma P, Sahbaee P, Schoepf UJ, Ravenel JG, et al. Deep learning to convert unstructured CT pulmonary angiography reports into structured reports. *Eur Radiol Exp.* (2019) 3:37. doi: 10.1186/s41747-019-0118-1
17. Piotrkowicz A, Johnson O, Hall G. Finding relevant free-text radiology reports at scale with IBM Watson content analytics: a feasibility study in the UK NHS. *J Biomed Semantics.* (2019) 10:21. doi: 10.1186/s13326-019-0213-5
18. Banerjee I, Bozkurt S, Caswell-Jin JL, Kurian AW, Rubin DL. Natural language processing approaches to detect the timeline of metastatic recurrence of breast cancer. *JCO Clin Cancer Inform.* (2019) 3:1–12. doi: 10.1200/CCI.19.00034
19. Bogaerts J, Ford R, Sargent D, Schwartz LH, Rubinstein L, Lacombe D, et al. Individual patient data analysis to assess modifications to the RECIST criteria. *Eur J Cancer.* (2009) 45:248–60. doi: 10.1016/j.ejca.2008.10.027
20. Nguyen NC, Vercher-Conejero JL, Sattar A, Miller MA, Maniawski PJ, Jordan DW, et al. Image quality and diagnostic performance of a digital PET prototype in patients with oncologic diseases: initial experience and comparison with analog PET. *J Nucl Med.* (2015) 56:1378–85. doi: 10.2967/jnumed.114.148338
21. Hirata K, Kobayashi K, Wong KP, Manabe O, Surmak A, Tamaki N, et al. A semi-automated technique determining the liver standardized uptake value reference for tumor delineation in FDG PET-CT. *PLoS ONE.* (2014) 9:e105682. doi: 10.1371/journal.pone.0105682
22. FDG-PET/CT Technical Committee. *FDG-PET/CT as an Imaging Biomarker Measuring Response to Cancer Therapy, Quantitative Imaging Biomarkers Alliance, Version 1.13*. Technically Confirmed Version. Available online at: http://qibawiki.rsna.org/images/1/1f/QIBA_FDG-PET_Profile_v113.pdf. *RSNAORG/QIBA* (accessed November 18, 2016).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 18 June 2021
doi: 10.3389/fmed.2021.673876

Pre-Operative Prediction of Mediastinal Node Metastasis Using Radiomics Model Based on ^{18}F-FDG PET/CT of the Primary Tumor in Non-Small Cell Lung Cancer Patients

Kai Zheng[1,2,3], *Xinrong Wang*[4], *Chengzhi Jiang*[2], *Yongxiang Tang*[1], *Zhihui Fang*[1], *Jiale Hou*[1], *Zehua Zhu*[1] *and Shuo Hu*[1,5,6*]

[1] *Department of Nuclear Medicine, Xiangya Hospital, Central South University, Changsha, China,* [2] *Positron Emission Tomography/Computed Tomography (PET/CT) Center, Hunan Cancer Hospital, Changsha, China,* [3] *The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, China,* [4] *General Electric (GE) Healthcare (China), Shanghai, China,* [5] *National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China,* [6] *Key Laboratory of Biological Nanotechnology of National Health Commission, Xiangya Hospital, Central South University, Changsha, China*

OPEN ACCESS

Edited by:
Xiaoli Lan,
Huazhong University of Science and Technology, China

Reviewed by:
Salvatore Annunziata,
Catholic University of the Sacred Heart, Italy
Lidia Strigari,
Regina Elena National Cancer Institute (IRCCS), Italy

****Correspondence:***
Shuo Hu
hushuo2018@163.com

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: *28 February 2021*
Accepted: *11 May 2021*
Published: *18 June 2021*

Citation:
Zheng K, Wang X, Jiang C, Tang Y, Fang Z, Hou J, Zhu Z and Hu S (2021) Pre-Operative Prediction of Mediastinal Node Metastasis Using Radiomics Model Based on ^{18}F-FDG PET/CT of the Primary Tumor in Non-Small Cell Lung Cancer Patients. Front. Med. 8:673876. doi: 10.3389/fmed.2021.673876

Purpose: We investigated whether a fluorine-18-fluorodeoxy glucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT)-based radiomics model (RM) could predict the pathological mediastinal lymph node staging (pN staging) in patients with non-small cell lung cancer (NSCLC) undergoing surgery.

Methods: A total of 716 patients with a clinicopathological diagnosis of NSCLC were included in this retrospective study. The prediction model was developed in a training cohort that consisted of 501 patients. Radiomics features were extracted from the ^{18}F-FDG PET/CT of the primary tumor. Support vector machine and extremely randomized trees were used to build the RM. Internal validation was assessed. An independent testing cohort contained the remaining 215 patients. The performances of the RM and clinical node staging (cN staging) in predicting pN staging (pN0 vs. pN1 and N2) were compared for each cohort. The area under the curve (AUC) of the receiver operating characteristic curve was applied to assess the model's performance.

Results: The AUC of the RM [0.81 (95% CI, 0.771–0.848); sensitivity: 0.794; specificity: 0.704] for the predictive performance of pN1 and N2 was significantly better than that of cN in the training cohort [0.685 (95% CI, 0.644–0.728); sensitivity: 0.804; specificity: 0.568], (P-value = 8.29e-07, as assessed by the Delong test). In the testing cohort, the AUC of the RM [0.766 (95% CI, 0.702–0.830); sensitivity: 0.688; specificity: 0.704] was also significantly higher than that of cN [0.685 (95% CI, 0.619–0.747); sensitivity: 0.799; specificity: 0.568], (P = 0.0371, Delong test).

Conclusions: The RM based on ^{18}F-FDG PET/CT has a potential for the pN staging in patients with NSCLC, suggesting that therapeutic planning could be tailored according to the predictions.

Keywords: non-small cell lung cancer, ^{18}F-FDG PET/CT, radiomics analysis, lymph node staging, predict, primary tumor

INTRODUCTION

Among all cancers, lung cancer remains the most commonly diagnosed (11.6% of the total cases) and leading cause of cancer death (18.4% of the total cancer deaths). Non-small cell lung cancer (NSCLC) accounts for 85% of the cases (1, 2). For patients newly diagnosed with NSCLC, the exact evaluation of the pathological lymph node (LN) status plays an important role in the choice of therapy regimen. There is a consensus that lobectomy combined with systemic nodal dissection is the recommended surgical treatment for early-stage NSCLC; however, sublobar resection and stereotactic body radiation therapy (SBRT) are possible alternatives for patients who are ineligible for lobectomy (3–5). Thus, accurate differentiation of pathological node-negative from positive is critical for selecting the optimal therapeutic plan.

Currently, one of the most widespread modalities used for the clinical LN (cN) staging of patients with NSCLC is fluorine-18-fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) (6, 7). Unfortunately, the accuracy of ^{18}F-FDG PET for the direct evaluation of each mediastinal LN for the presence of metastasis is inherently limited by an avid FDG uptake that can be caused by inflammation due to infectious or non-infectious etiologies such as tuberculosis, pneumoconiosis, or chronic obstructive pulmonary disease (8–10). To improve the diagnostic ability of false-positive signs, several studies have analyzed the differences in parameters such as morphology, density, metabolism, and radiomics between benign and malignant LNs (11, 12). Nevertheless, because of the low FDG uptake, occult LN metastasis (OLM) in patients with NSCLC fails to be detected by ^{18}F-FDG PET (13) and, hence, imaging is prone to false-negative signs. Accordingly, some researchers have tried to predict OLM by analyzing the ^{18}F-FDG metabolic parameters of the primary tumor in NSCLC (14, 15). To the best of our knowledge, few researchers have dealt with both problems of false-positive signs and OLM in mediastinal LN staging. Furthermore, few studies have investigated whether radiomics features derived from the primary lesion of NSCLC might provide useful information for mediastinal LN staging.

Therefore, we constructed and validated a radiomics model (RM) to predict the pathological mediastinal LN staging (pN0 vs. pN1 and pN2) based on the ^{18}F-FDG PET/CT imaging of NSCLC primary tumors.

MATERIALS AND METHODS

Patients

This study involving human participants was reviewed and approved by the Ethical Commission of Medical Research Involving Human Subjects at the Region of Xiangya Hospital, Central South University, China, and the requirement for informed consent was waived. We reviewed the electronic medical records of 716 consecutive patients with NSCLC [adenocarcinoma (ADC) and squamous cell carcinoma (SCC)] who underwent both ^{18}F-FDG PET/CT staging and surgical resection with a curative intent from February 2007 to November 2019. All the patients underwent surgical resection with systematic mediastinal (N2) and hilar (N1) LN dissections within 2 weeks of ^{18}F-FDG PET/CT examination. Pre-operative cN staging and post-operative pN staging of the patients were performed and recorded according to the eighth edition of the Union for International Cancer Control TNM classification (16). Histological types were diagnosed according to the World Health Organization classification. We excluded patients from the study if they had (i) histology other than ADC or SCC, (ii) history of other cancer, (iii) received any treatment before ^{18}F-FDG PET/CT, and (iv) undergone pre-operative lung biopsy.

^{18}F-FDG PET/CT Acquisition and Reconstruction

All ^{18}F-FDG PET/CT scans were performed on a dedicated PET/CT scanner (Discovery ST8, GE Healthcare, Chicago, IL). All patients fasted for at least 6 h before imaging, and a blood glucose level of $<$110 mg/dL was confirmed before the administration of ^{18}F-FDG. PET/CT was performed $\sim$60 min after the intravenous injection of 370 MBq/kg of ^{18}F-FDG. First, a low-dose CT scan without contrast enhancement (120 mA, 150 kV, 512 $\times$ 512 matrix, the pitch of 1.75, reconstruction thickness and interval of 3.75 mm) for a precise anatomical localization and attenuation correction was performed. Next, a three-dimensional PET scan (thickness of 3.27 mm) was performed from the skull base to the proximal thighs with an acquisition time of 3 min per bed position.

The PET data sets were iteratively reconstructed using an ordered-subset expectation maximization (OSEM) algorithm with attenuation correction. All collected images were displayed on the GE Healthcare Xeleris 3.0 to reconstruct the PET, CT, and PET/CT fusion images.

Image Interpretation and Lesion Segment

Two experienced nuclear medicine physicians who were blinded to the patient's clinical information retrospectively reviewed the ^{18}F-FDG PET/CT scans. Any difference of opinion was resolved by consensus. Mediastinal and hilar LNs with a short axis of $\geq$10 mm in the short axis on CT and with a high accumulation of ^{18}F-FDG compared with that of the adjacent mediastinal tissue were considered as cN2 or cN1 at our institution. Fused PET/CT images were viewed on the Advantage Workstation (version AW 4.7, GE Healthcare).

The region of interest (ROI) for each patient was delineated initially around the tumor outline for the largest cross-sectional area of the primary lung lesion on both the CT and PET images. The ROIs were segmented manually by a single experienced nuclear medicine physician, and the final ROIs were checked by another nuclear medicine physician with more than 10 years of experience in PET/CT diagnosis. The open-source imaging platform ITK-SNAP software (version 3.6; www.itksnap.org) was used to plot the ROIs of the corresponding lesions (17). The feature data were extracted, pre-processed, modeled, evaluated, and validated using the scikit-learn (sklearn, scikit-learn.org) packages in the python platform (18).

Radiomics Feature Extraction

Data pre-processing: to ensure that the features were comparable, training/testing cohort division, missing-value filling, and data standardization were performed. First, to maintain the distribution of the original data, a stratified sampling method was applied to identify the training (501 samples, 70%) and testing cohorts (215 samples, 30%) (19). Moreover, the missing values (0 and 5) were filled with the median in the training and testing cohorts, respectively, and then, the same normalization was used for the data.

There were 1,438 features of primary tumors that were automatically extracted using the sklearn packages. The Spearman rank order correlation coefficient was used to calculate the relationship between features, and the redundant features were eliminated with an average absolute correlation of 0.85 as the threshold. Support vector machine–recursive feature elimination, and the extremely randomized trees were applied to reduce the dimensions and select optimized features for the radiomics model (RM) to avoid the impact of redundant and unconnected features. The relevance of the association between each radiomics feature was established using heat maps (**Figure 1**). Consequently, a total of 25 principal correlative features, obtained through dimension reduction, were identified for inclusion in the RM to distinguish pN0 from pN1 and pN2. The results of the feature selection are shown in **Table 1**.

Radiomics Modeling and Evaluation

Extremely randomized trees was used as a classifier to model and optimize the radiomics signature in the modeling process. The 25 selected features were put into the classifier to build the RM to predict the pathological status of mediastinal LNs in the training cohort. Thereafter, a five-fold cross-validation of the training cohort was used to identify differences in the results. The training model was applied to the testing cohort for model validation. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve was used as a means of quantitatively identifying the effective performance of the RM. The confusion matrix in the testing cohort was calculated (**Figure 2**).

Statistical Analysis

Statistical analyses were conducted using the SPSS software, version 23.0 (IBM Corm., Armonk, NY), and P-values < 0.05 were deemed statistically significant.

The predictive abilities of the RM and cN were investigated using ROC analysis. The statistical significance of the improvement in the AUC after adding an explanatory factor was evaluated using the Delong test (20).

The clinicopathologic characteristics of the patients with a pN0 status were compared with those of the patients with pN1 and pN2. The training and testing data cohorts were compared to identify factors contributing to nodal metastasis

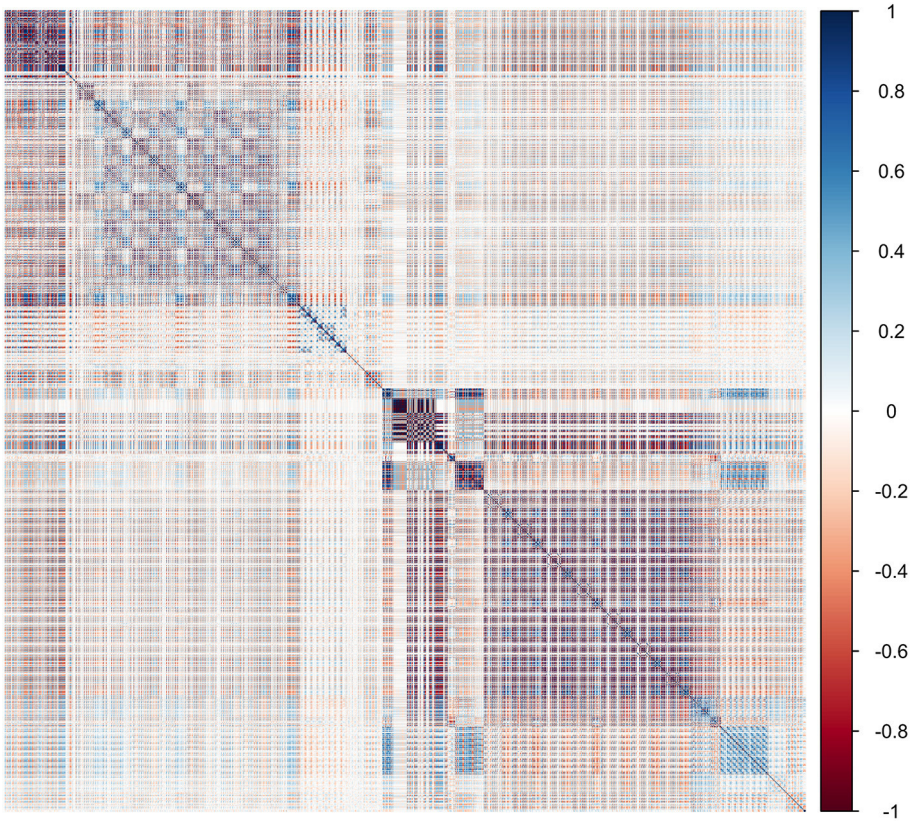


FIGURE 1 | Heat map showing the correlation of radiomics features in the training cohort. The intensity of the relevance of each feature is displayed as a certain color. The darker the color, the higher the relevance, and the lighter the color, the lower the relevance.

TABLE 1 | The list of selected radiomics features.

Characteristic type	Description	Selected features
Histogram feature	Histogram parameters are related to the properties of individual pixels. They describe the distribution of voxel intensities within the images through the commonly used and basic metrics. Let X denote the 3D image matrix with voxels and the first-order histogram divided by discrete intensity levels.	PET_original_firstorder_Minimum
Textural phenotype features	Texture is one of the important characteristics used in identifying objects or regions of interest in an image. Texture represents the appearance of the surface and how its elements are distributed. It is considered an important concept in machine vision; in a sense, it assists in predicting the feeling of the surface (e.g., smoothness, coarseness, etc.) from image.	PET_textural_phenotype_level_H
Intra-peri-nodular textural transition features	Intra-peri-nodular textural transition features represents a minimal set of quantitative measurements which attempt to capture the transitional heterogeneity from the intra- to the peri-nodular space.	PET_lpris_shell0_ge_mean
Partial local pattern binary feature	Partial local pattern binary feature is a local descriptor of the image based on the neighborhood for any given pixel. The neighborhood of a pixel is given in the form of P number of neighbors within a radius of R.	PET_PLBP_hist_tumor_orient6_0 CT_PLBP_hist_tumor_orient2_7 CT_PLBP_hist_tumor_orient2_3 PET_PLBP_hist_tumor_orient3_1 PET_PLBP_hist_tumor_orient4_3 CT_PLBP_hist_tumor_orient1_2
High order texture feature based on wavelet transform	By using a family of functions localized in terms of time and frequency, wavelet transforms can centralize the energy of the original image within only a few coefficients after wavelet decomposition. These coefficients have high local relativity in three directions of different sub-band images: horizontal, vertical, and diagonal.	CT_wavelet-LHL_lbp-3D-m2_firstorder_90Percentile CT_wavelet-LLL_lbp-3D-m2_firstorder_InterquartileRange PET_wavelet-HLL_lbp-3D-m2_firstorder_Median PET_wavelet-HHL_lbp-3D-m1_firstorder_Skewness CT_wavelet-LHH_lbp-3D-m1_firstorder_Median PET_wavelet-LHL_lbp-3D-m1_firstorder_Median CT_wavelet-HLL_lbp-3D-m1_firstorder_90Percentile CT_WL_lbp_hist_cH1_1 PET_WL_lbp_hist_cD1_4 PET_wavelet-HLL_lbp-3D-m1_firstorder_Median CT_wavelet-HLL_lbp-3D-m2_firstorder_Range PET_wavelet-HHL_lbp-3D-k_firstorder_Minimum PET_WL_lbp_hist_cH2_2 CT_wavelet-LLL_lbp-3D-m1_firstorder_Median CT_WL_lbp_hist_cV2_7

using the χ^2-test for categorical data and the one-sample *t*-test for continuous variables.

RESULTS

Characteristics of All Patients

The clinicopathological characteristics of the 716 patients enrolled in the study are shown in **Table 2**. **Figure 3** shows the patient recruitment pathway. Among them, 220 were female and 496 were male, with an age range from 25 to 78 years. ADC was the most common histological type of NSCLC (417/716). The number of patients with SCC was 329. In the training cohort, the number of patients with pN0, pN1, and pN2 were 315, 74, and 112, respectively. In the testing cohort, the number of patients with pN0, pN1, and pN2 were 135, 43, and 37, respectively. The age of the training cohort, sex of the testing cohort, and

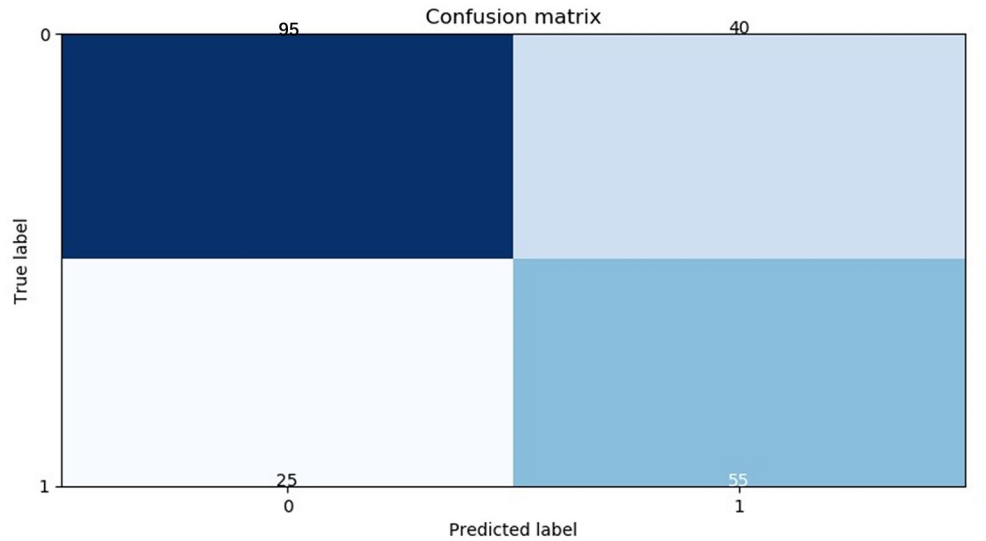


FIGURE 2 | Confusion matrix of the radiomics model in the testing cohort. The abscissas and ordinates represent the true and predictive labels, respectively.

TABLE 2 | Clinical characteristics.

Characteristic	Training cohort		*P*	Testing cohort		*P*
	pN0 (*n* = 315)	pN1&2 (*n* = 186)		pN0 (*n* = 135)	pN1&2 (*n* = 80)	
Age, mean ± SD, years	60 ± 9	58 ± 9	0.005[a]	60 ± 9	58 ± 9	0.152[a]
Gender, No. (%)			0.627			0.041
Male	212	128		90	66	
Female	103	58		45	14	
Smoking history			0.810			0.206
Yes	181	108		78	52	
No	134	78		57	28	
Lobar distribution			0.214			0.385
LUL	77	45		40	25	
LLL	46	28		25	15	
RUL	57	52		17	21	
RML	28	13		11	1	
RLL	107	48		42	18	
Anatomical classification			0.02			0.008
Central lung cancer	53	52		25	27	
Peripheral lung cancer	263	133		110	53	
Histologic cell type			0.696			0.697
SCC	109	67		52	38	
ADC	207	119		83	42	

RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe.
[a]*one sample T-test.*

anatomical classification of the two cohorts were statistically significantly different between the pN0 status and pN1 and pN2 status ($P < 0.050$). However, no significant difference was observed in the age of the testing cohort and in the sex of the training cohort ($P > 0.050$). Moreover, smoking history, lobar distribution, and histologic cell type were not significantly different between the two cohorts.

RM Performance

The diagnostic efficiency of the RM and cN were evaluated by the ROC curve. The AUC of the RM [0.81 (95% CI, 0.771–0.848); sensitivity: 0.794; specificity: 0.704] for the predictive performance of the pathological node status was significantly better than that of the cN in the training cohort [0.685 (95% CI, 0.644–0.728); sensitivity: 0.804;

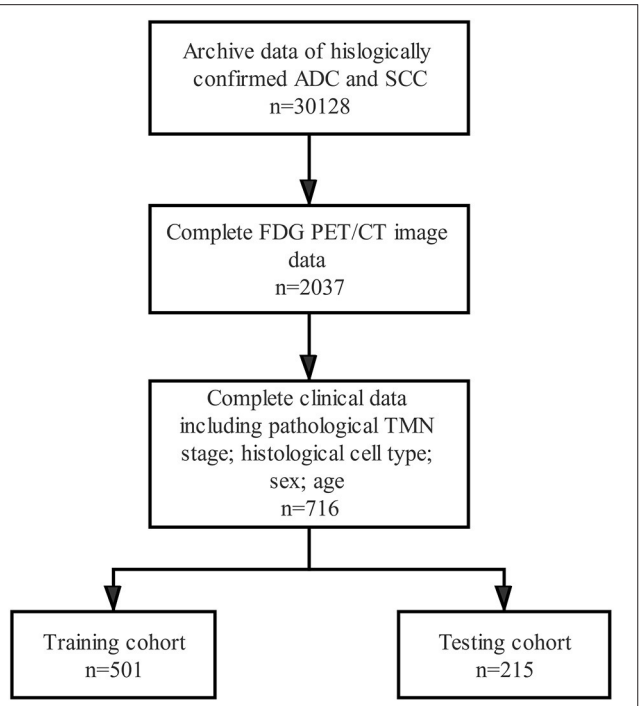


FIGURE 3 | Flow diagram shows patient selection details.

specificity: 0.568], (P = 8.29e-07, as assessed using the Delong test).

In the testing cohort, the AUC of the RM [0.766 (95% CI, 0.702–0.830); sensitivity: 0.688; specificity: 0.704] was also significantly higher than that of the cN [0.685 (95% CI, 0.619–0.747); sensitivity: 0.799; specificity: 0.568], (P = 0.0371, as assessed using the Delong test).

The above-mentioned nuclear medicine physicians excluded the patients with LNs significant enlargement and intense ^{18}F-FDG uptake in PET/CT, and confirmed N1 or N2 by pathology from the whole population. The remaining 634 patients were defined as the cN ± group. Then, the sensitivity, specificity, and AUC of the cN and the RM in the cN ± group were calculated, respectively.

The RM showed the AUC of 0.802 (95%CI, 0.683–0.921) for the prediction of mediastinal LNs malignancy using the optimum cutoff value of 0.382 in the cN ± group. The sensitivity and specificity of the RM were 0.718 and 0.767, respectively. In comparison, the cN showed the AUC of 0.611 (95%CI, 0.572–0.650) using the optimum cutoff value of 0.500. The sensitivity and specificity of the cN were 0.819 and 0.404, respectively. The results demonstrated that the performance of the RM was effective in discriminating pN0 from pN1 and pN2 using the ^{18}F-FDG PET/CT images. The performances of the RM and cN in the training and testing cohorts and the cN ± group are displayed in detail in **Figures 4**, **5**, respectively, and the representative cases are presented in **Figure 6**.

DISCUSSION

In cases of NSCLC with a chance of cure, the standard surgical procedure is a pulmonary lobectomy with systemic mediastinal nodal dissection. However, some patients are not eligible for this therapy because of their advanced age or the presence of severe medical diseases, and some patients refuse surgical treatment. Limited surgery (wedge resection or segmentectomy) or SBRT would be alternatives for such patients. Sublobar resection helps preserve more healthy lung tissue, shortens the operative time, and improves the post-operative quality of life. Perioperative mortality and operative complication morbidity do not differ significantly between lobar and sublobar resection (3). SBRT has emerged as the preferred management strategy for patients who are not surgical candidates; however, for the selection of SBRT or restrictive surgery, accurate prediction of a pathological LN-negative status is a pre-requisite.

The diagnosis of NSCLC mediastinal LN metastasis is generally based on several parameters such as metabolism, size, morphology, and attenuation, which leads to dependence on clinical experience. In other words, the traditional practice involves treating medical images as pictures intended solely for visual interpretation (21). False-positive findings of mediastinal LNs are not uncommon in functional imaging with ^{18}F-FDG PET/CT because the modality can mistakenly identify inflammation in patients with NSCLC due to infection, inflammation, or granulomatous diseases (8–10). The main molecular and pathological mechanisms of an avid FDG uptake in benign mediastinal LNs are lymphoid follicular hyperplasia and histiocyte infiltration associated with glucose transporter-1 overexpression (22). When benign mediastinal LNs manifest as a false-positive finding on PET imaging, and the CT morphology is not informative enough to support a judgment, there is an increased risk of an incorrect diagnosis. Benign high-uptake LNs can coexist with occult metastasis, making an accurate cN staging more difficult. In the clinical practice of mediastinal LN staging in NSCLC, nuclear medicine physicians are faced with the challenge of suspected positive LNs and possible OLM almost every day, which is difficult to deal with by relying solely on experience.

In the existing studies, for the accuracy of the cN staging in NSCLC, radiologists and nuclear physicians often analyzed parameters such as morphology and glucose metabolism or radiomic features of visible mediastinal LNs to improve the diagnostic ability of ^{18}F-FDG PET/CT for metastasis (23–25). Gao et al. researched the method and efficacy of support vector machine classifiers based on texture features and a multi-resolution histogram to evaluate mediastinal LNs (11). Flechsig et al. used density as a threshold for the detection of malignant LN infiltration in a radiomics analysis of patients with NSCLC (12). Likewise, Lee et al.'s research indicated that the risk of mediastinal LN metastasis in NSCLC patients could be further stratified using both ^{18}F-FDG uptake and LN density (24). Cho et al. attempted to determine the optimal cut-off values of the mediastinal LN standardized uptake values (SUV-LN)/primary tumor SUV (SUV-T) ratio to discriminate metastatic LNs from benign LNs (26). However, these researchers faced the common problem of an unpredictable OLM. Some algorithms for the

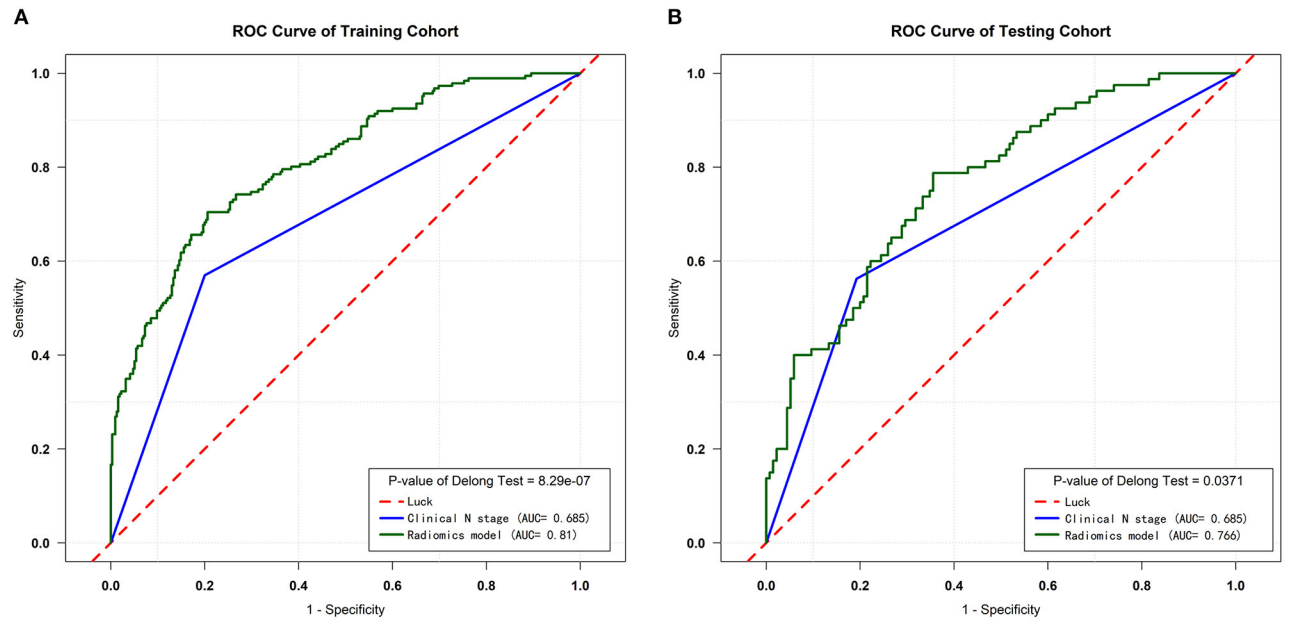


FIGURE 4 | ROC curves of the training and testing cohorts. **(A)** ROC curves for the RM and cN of the training cohort. **(B)** ROC curves for the RM and cN of the testing cohort.

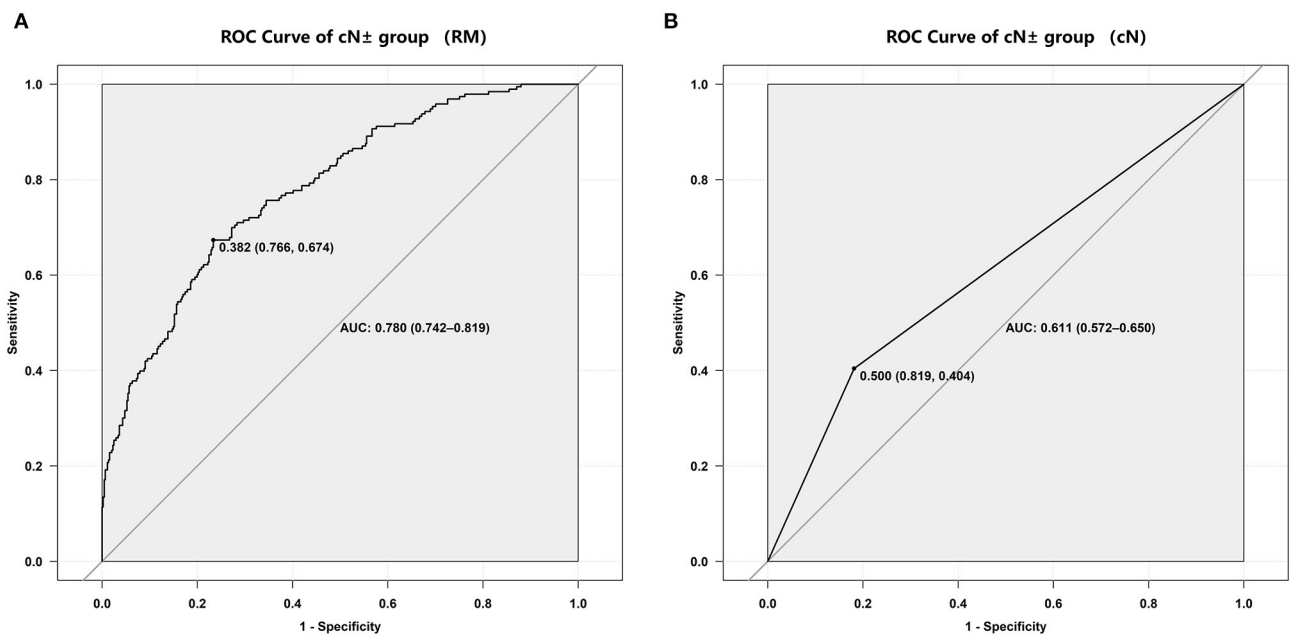


FIGURE 5 | ROC curves of the cN ± group. **(A)** ROC curve for the RM. **(B)** ROC curve for the cN.

analysis of parameters based on ^{18}F-FDG uptake have been proposed in light of these limitations. Ouyang et al. used the primary tumor-to-blood SUV ratio and metabolic parameters in clinical N0 lung ADC to predict OLM (14). Kim et al. investigated the OLM's predictability using SUV, metabolic tumor volume (MTV), and total lesion glycolysis (TLG) in patients with cN0 lung SCC before surgery (15). There are no current studies of ^{18}F-FDG PET/CT primary tumor-based

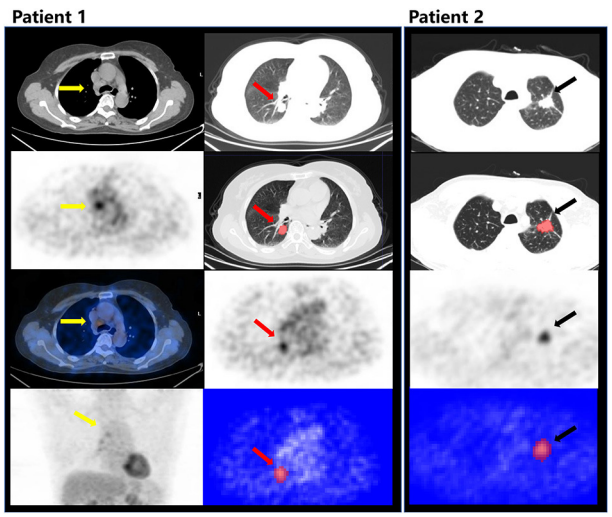


FIGURE 6 | Patient 1: male, 55 years old, ADC, PET/CT showed lymph nodes in 4R (yellow arrow) with an intense FDG uptake, which was evaluated as cN2, predicted as N0 by the RM, and confirmed as pN0 after radical resection. The red arrow indicates the tumor in the right lower lobe (CT, PET) and the segmentation on ITK-SNAP. Patient 2: male, 66 years old, ADC, evaluated as cN0, but predicted as N+ by the RM, and confirmed as pN1 after radical resection. The black arrow indicates the primary lesion in the left upper lobe (CT, PET) and the segmentation on ITK-SNAP.

radiomics classifiers of the LN staging (N0 vs. N1 and N2) in NSCLC. Therefore, if our solution proves to be feasible, it can be used to either differentiate benign and malignant LNs or determine OLM, thus leading to an informed therapeutic decision-making in the face of the challenge of false-positive and false-negative images.

Huang et al. have developed and validated a radiomics nomogram based on the primary tumor in a contrast-enhanced CT for pre-operative LN metastasis prediction in colorectal cancer (27). Inspired by their research achievement, we aim to introduce the radiomics modeling approach based on the ^{18}F-FDG PET-CT images of the primary lesion into the LN staging in NSCLC. On the basis of the radiomics hypothesis, intratumoral heterogeneity detected by imaging could be the expression of genomic heterogeneity, which implies a worse prognosis because tumors with more genomic heterogeneity are more likely to be resistant to treatment metastasis (28). Mediastinal LN staging in NSCLC is highly correlated with prognosis; therefore, we assumed that LN metastasis information may be obtained from intratumoral heterogeneity. Some studies had discovered pre-therapy ^{18}F-FDG PET/CT or CT-based radiomics classifiers of survival or response in patients with NSCLC (29–32). Therefore, it can obtain information from the primary lesion that is helpful for diagnosis or prognosis.

Nevertheless, the present study has several limitations too. Due to its retrospective design and performance at a single center, there is a risk of selection bias. A larger, multi-institutional prospective randomized study is needed to confirm these results.

CONCLUSIONS

A Radiomics Model based on the ^{18}F-FDG PET/CT analysis provided useful information for mediastinal LN staging in patients with NSCLC. Therefore, therapeutic planning could be tailored according to predictions, and limited surgery or SBRT could be helpful in patients with cN0.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Commission of Medical Research Involving Human Subjects at Region of Xiangya Hospital, Central South University, China. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

KZ designed the method, acquired the data, and wrote and edited the article. SH designed the method and approved the manuscript. Image analysis was performed by XW. Statistical analysis was performed by XW and CJ. YT, ZF, JH, and ZZ aided in data acquisition and interpretation. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (Nos. 91859207, 81771873, and 81471689).

ACKNOWLEDGMENTS

We thank the patients who were included in this research project and the staff who developed the ITK-SNAP and sklearn package.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Erratum: global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2020) 70:313. doi: 10.3322/caac.21609
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. (2016) 66:115–32. doi: 10.3322/caac.21338

3. Altorki NK, Wang X, Wigle D, Gu L, Darling G, Ashrafi AS, et al. Perioperative mortality and morbidity after sublobar versus lobar resection for early-stage non-small-cell lung cancer: *post-hoc* analysis of an international, randomised, phase 3 trial (CALGB/Alliance 140503). *Lancet Respir Med*. (2018) 6:915–24. doi: 10.1016/S2213-2600(18)30411-9
4. Ball D, Mai GT, Vinod S, Babington S, Ruben J, Kron T, et al. Stereotactic ablative radiotherapy versus standard radiotherapy in stage 1 non-small-cell lung cancer (TROG 09.02 CHISEL): a phase 3, open-label, randomised controlled trial. *Lancet Oncol*. (2019) 20:494–503. doi: 10.1016/S1470-2045(18)30896-9
5. Phillips I, Sandhu S, Luchtenborg M, Harden S. Stereotactic ablative body radiotherapy versus radical radiotherapy: comparing real-world outcomes in stage i lung cancer. *Clin Oncol (R Coll Radiol)*. (2019) 31:681–7. doi: 10.1016/j.clon.2019.07.013
6. Abramyuk A, Appold S, Zophel K, Hietschold V, Baumann M, Abolmaali N. Quantitative modifications of TNM staging, clinical staging and therapeutic intent by FDG-PET/CT in patients with non small cell lung cancer scheduled for radiotherapy–a retrospective study. *Lung Cancer*. (2012) 78:148–52. doi: 10.1016/j.lungcan.2012.08.001
7. Rohren EM, Turkington TG, Coleman RE. Clinical applications of PET in oncology. *Radiology*. (2004) 231:305–32. doi: 10.1148/radiol.2312021185
8. Kang F, Wang S, Tian F, Zhao M, Zhang M, Wang Z, et al. Comparing the diagnostic potential of 68Ga-alfatide II and 18F-FDG in differentiating between non-small cell lung cancer and tuberculosis. *J Nucl Med*. (2016) 57:672–7. doi: 10.2967/jnumed.115.167924
9. Choi EK, Park HL, Yoo IR, Kim SJ, Kim YK. The clinical value of F-18 FDG PET/CT in differentiating malignant from benign lesions in pneumoconiosis patients. *Eur Radiol*. (2020) 30:442–51. doi: 10.1007/s00330-019-06342-1
10. Flechsig P, Kratochwil C, Schwartz LH, Rath D, Moltz J, Antoch G, et al. Quantitative volumetric CT-histogram analysis in N-staging of 18F-FDG-equivocal patients with lung cancer. *J Nucl Med*. (2014) 55:559–64. doi: 10.2967/jnumed.113.128504
11. Gao X, Chu C, Li Y, Lu P, Wang W, Liu W, et al. The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from (18)F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer. *Eur J Radiol*. (2015) 84:312–7. doi: 10.1016/j.ejrad.2014.11.006
12. Flechsig P, Frank P, Kratochwil C, Antoch G, Rath D, Moltz J, et al. Radiomic analysis using density threshold for FDG-PET/CT-based N-staging in lung cancer patients. *Mol Imaging Biol*. (2017) 19:315–22. doi: 10.1007/s11307-016-0996-z
13. Farjah F, Lou F, Sima C, Rusch VW, Rizk NP. A prediction model for pathologic N2 disease in lung cancer patients with a negative mediastinum by positron emission tomography. *J Thorac Oncol*. (2013) 8:1170–80. doi: 10.1097/JTO.0b013e3182992421
14. Ouyang ML, Tang K, Xu MM, Lin J, Li TC, Zheng XW. Prediction of occult lymph node metastasis using tumor-to-blood standardized uptake ratio and metabolic parameters in clinical N0 lung adenocarcinoma. *Clin Nucl Med*. (2018) 43:715–20. doi: 10.1097/RLU.0000000000002229
15. Kim DH, Song BI, Hong CM, Jeong SY, Lee SW, Lee J, et al. Metabolic parameters using (1)(8)F-FDG PET/CT correlate with occult lymph node metastasis in squamous cell lung carcinoma. *Eur J Nucl Med Mol Imaging*. (2014) 41:2051–7. doi: 10.1007/s00259-014-2831-6
16. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The eighth edition lung cancer stage classification. *Chest*. (2017) 151:193–203. doi: 10.1016/j.chest.2016.10.010
17. Yushkevich PA, Gerig G. ITK-SNAP: an intractive medical image segmentation tool to meet the need for expert-guided segmentation of complex medical images. *IEEE Pulse*. (2017) 8:54–7. doi: 10.1109/MPUL.2017.2701493
18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. (2011) 12:2825–30.
19. Zhao C-K, Ren T-T, Yin Y-F, Shi H, Wang H-X, Zhou B-Y, et al. A comparative analysis of two machine learning-based diagnostic patterns with thyroid imaging reporting and data system for thyroid nodules: diagnostic performance and unnecessary biopsy rate. *Thyroid*. (2021) 31:470–81. doi: 10.1089/thy.2020.0305
20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. (1988) 44:837–45. doi: 10.2307/2531595
21. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169
22. Kwon SY, Min JJ, Song HC, Choi C, Na KJ, Bom HS. Impact of lymphoid follicles and histiocytes on the false-positive FDG uptake of lymph nodes in non-small cell lung cancer. *Nucl Med Mol Imaging*. (2011) 45:185–91. doi: 10.1007/s13139-011-0085-9
23. Giesel FL, Schneider F, Kratochwil C, Rath D, Moltz J, Holland-Letz T, et al. Correlation between SUVmax and CT radiomic analysis using lymph node density in PET/CT-based lymph node staging. *J Nucl Med*. (2017) 58:282–7. doi: 10.2967/jnumed.116.179648
24. Lee JW, Kim EY, Kim DJ, Lee JH, Kang WJ, Lee JD, et al. The diagnostic ability of (18)F-FDG PET/CT for mediastinal lymph node staging using (18)F-FDG uptake and volumetric CT histogram analysis in non-small cell lung cancer. *Eur Radiol*. (2016) 26:4515–23. doi: 10.1007/s00330-016-4292-8
25. Yin G, Song Y, Li X, Zhu L, Su Q, Dai D, et al. Prediction of mediastinal lymph node metastasis based on (18)F-FDG PET/CT imaging using support vector machine in non-small cell lung cancer. *Eur Radiol*. (2020) 31:3983–92. doi: 10.1007/s00330-020-07466-5
26. Cho J, Choe JG, Pahk K, Choi S, Kwon HR, Eo JS, et al. Ratio of mediastinal lymph node SUV to primary tumor SUV in (18)F-FDG PET/CT for nodal staging in non-small-cell lung cancer. *Nucl Med Mol Imaging*. (2017) 51:140–6. doi: 10.1007/s13139-016-0447-4
27. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
28. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*. (2010) 467:1109–13. doi: 10.1038/nature09460
29. Arshad MA, Thornton A, Lu H, Tam H, Wallitt K, Rodgers N, et al. Discovery of pre-therapy 2-deoxy-2-18F-fluoro-D-glucose positron emission tomography-based radiomics classifiers of survival outcome in non-small-cell lung cancer patients. *Eur J Nucl Med Mol Imaging*. (2018) 46:455–66. doi: 10.1007/s00259-018-4139-4
30. Kirienko M, Cozzi L, Antunovic L, Lozza L, Fogliata A, Voulaz E, et al. Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery. *Eur J Nucl Med Mol Imaging*. (2018) 45:207–17. doi: 10.1007/s00259-017-3837-7
31. Dong X, Sun X, Sun L, Maxim PG, Xing L, Huang Y, et al. Early change in metabolic tumor heterogeneity during chemoradiotherapy and its prognostic value for patients with locally advanced non-small cell lung cancer. *PLoS ONE*. (2016) 11:e0157836. doi: 10.1371/journal.pone.0157836
32. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology*. (2016) 281:947–57. doi: 10.1148/radiol.2016152234

Conflict of Interest: XW was employed by the company GE healthcare (China).

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 05 July 2021
doi: 10.3389/fmed.2021.646862

^{18}F-PEG1-Vinyl Sulfone-Labeled Red Blood Cells as Positron Emission Tomography Agent to Image Intra-Abdominal Bleeding

Xinyi Zhang[1,2,3†], *Li Wang*[1,2,3†], *Wenhui Fu*[1,2,3], *Yue Feng*[1,2,3], *Chengrun Zeng*[1,2,3], *Liu Zhou*[1,2,3,4], *Tao Zhang*[5], *Tingting Xu*[1,2,3], *Jianpeng Cao*[1,2,3], *Zibo Li*[5*] and *Yue Chen*[1,2,3*]

[1] Department of Nuclear Medicine, The Affiliated Hospital of Southwest Medical University, Luzhou, China, [2] Nuclear Medicine and Molecular Imaging Key Laboratory of Sichuan, Luzhou, China, [3] Academician (Expert) Workstation of Sichuan, Luzhou, China, [4] School of Pharmacy, Southwest Medical University, Luzhou, China, [5] Department of Radiology, Lineberger Comprehensive Cancer Center, and Biomedical Research Imaging Center, University of North Carolina, Chapel Hill, NC, United States

OPEN ACCESS

Edited by:
Xiaoli Lan,
Huazhong University of Science and Technology, China

Reviewed by:
Shuang Liu,
Purdue University, United States
Dawei Jiang,
Huazhong University of Science and Technology, China

***Correspondence:**
Zibo Li
zibo_li@med.unc.edu
Yue Chen
chenyue5523@126.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: 28 December 2020
Accepted: 02 June 2021
Published: 05 July 2021

Citation:
Zhang X, Wang L, Fu W, Feng Y, Zeng C, Zhou L, Zhang T, Xu T, Cao J, Li Z and Chen Y (2021) ^{18}F-PEG1-Vinyl Sulfone-Labeled Red Blood Cells as Positron Emission Tomography Agent to Image Intra-Abdominal Bleeding. Front. Med. 8:646862. doi: 10.3389/fmed.2021.646862

^{18}F-Labeled blood pool agents (BPAs) have attracted great attention for identifying bleeding sites. However, many BPAs are not sufficiently evaluated partially due to the limitations of labeling methods. In our previous work, we noticed that ^{18}F-PEG1-vinyl sulfone (^{18}F-VS) could efficiently label red blood cells (RBCs) *ex vivo* and *in situ*. However, its application as BPA is not fully evaluated. In this study, we systematically explored the feasibility of using ^{18}F-VS-labeled RBCs as a positron emission tomography (PET) BPA for intra-abdominal bleeding diagnosis. In brief, we first optimized the labeling conditions, which lead to an 80% labeling yield of RBCs after incubating with ^{18}F-VS in phosphate-buffered saline (PBS) at 37°C for 20 min. ^{18}F-VS-labeled RBCs were found to be stable *in vitro*, which could simplify its transportation/storage for *in vivo* applications. In normal rat PET study, the cardiovascular system could be clearly imaged up to 5 h post injection (p.i.). An intra-abdominal hemorrhage rat model demonstrated that the ^{18}F-VS-labeled RBCs clearly showed the dynamic changes of extravascular radioactivity due to intra-abdominal hemorrhage. Validation in the model of gastrointestinal bleeding clearly demonstrated the great potential of using ^{18}F-VS-labeled RBCs as a BPA, which could be further evaluated in future studies.

Keywords: ^{18}F-vinyl sulfone, red blood cell, blood pool imaging, intra-abdominal hemorrhage, positron emission tomography (PET)

INTRODUCTION

Blood pool imaging (BPI) is widely used in preclinical and clinical research including the detection of gastrointestinal bleeding (1), blood volume measurement (2), evaluation of cardiac function (3, 4), localization of hemangiomas (5), cerebral blood flow (6, 7), detection of infection (8), or lymphoma (9). Various kinds of BPI agents have been developed, including ^{18}F-FDG-, ^{111}In-oxine (8), or ^{99m}Tc-HMPAO-labeled leukocytes (10); ^{68}Ga-NOTA-NEB-(9) and ^{111}In-oxine-labeled platelets (11); ^{99m}Tc-PYP-labeled red blood cells (RBCs) (12); radionuclide-labeled peptides (13); and magnetic resonance angiography-based BPI agents (14). RBCs represent promising BPI agents

due to their good stability and easy availability. In fact, radionuclide-labeled RBCs have been used to obtain functional information of the cardiovascular system through quantitative analysis of BPI. During the past few decades, radionuclide-labeled RBCs [such as ^{51}Cr-RBCs (15), ^{99m}Tc-PYP-RBCs (16), and ^{68}Ga-oxine-RBCs (17)] have shown great progress in preclinical and clinical research applications. They have been used for cardiac function evaluation, diagnosis of hemangioma and digestive tract bleeding, cerebral blood flow measurement, and spleen imaging and positioning. Despite the progress, limitations of these BPI-agents include 1) unstable labeling yield: for example, the labeling rate of ^{99m}Tc-PYP-RBCs was greatly affected by the type and dose of the drug applied (18) and 2) the resolution of single-photon emission nuclide ^{99m}Tc often resulted in reduced image quality compared with the positron emission tomography (PET) nuclide (19). Therefore, researchers have been trying to develop BPI agents based on PET nuclides. ^{11}C, ^{13}N, and ^{68}Ga have been used to label RBCs, but their short half-lives or image resolution still limited their widespread clinical application to a certain extent (20, 21).

The positron nuclide ^{18}F has the advantages of good image quality and suitable half-life (109.8 min) for commercialization and transportation compared with other positron nuclides. Moreover, the resulting carbon–fluorine bonds generally have reasonable stability, which could be advantageous for BPI (22). However, there are only a few ^{18}F-labeled RBCs reported as BPI agents partially due to the limitation of the labeling method (23). Therefore, there is a need to find new methods that could lead to easily prepared ^{18}F-labeled BPI agents for preclinical and clinical diagnosis applications.

Previously, we established a new method for site-specific labeling of thiol groups based on ^{18}F-labeled vinyl sulfone (^{18}F-VS). Both peptides and proteins were found to react with ^{18}F-VS through the Michael addition in an aqueous system (24). Moreover, the resulting conjugates are stable in the aqueous solution and would not be hydrolyzed in a neutral solution like maleimide conjugates (25). Recently, it was also found that ^{18}F-VS could react with amino groups in addition to thiol groups even though the reaction is slower (26). Interestingly, we observed that ^{18}F-VS could efficiently label RBCs *in vitro* and *in vivo*. Despite the observation, the conditions to label RBCs were not optimized, and its application as BPI was not studied.

In this study, we evaluate the use of ^{18}F-((2-(2-fluoroethoxy)ethyl)sulfonyl)ethene (^{18}F-PEG1-vinyl sulfone [^{18}F-VS]) for RBC labeling, which are then applied as a new BPI agent for abdominal hemorrhage imaging in animal models.

MATERIALS AND METHODS

All chemicals involved in the synthesis were of reagent grade and purchased from Aladdin Bio-Chem Technology (Shanghai, China) or Sigma. The radiochemical purity was documented by high-performance liquid chromatography (LC-16). A gamma counter (CAPRAC-t, Huaruisen Technology Development Co., Ltd., Beijing, China) and a dose calibrator (CRC-15R, Capintec Inc., Florham Park, NJ) were used to measure the radioactivity of the samples. Mouse and rat data acquisitions were performed with a micro-PET/computed tomography (CT) scanner (Inveon, Siemens, Munich, Germany). Healthy Kunming mice (20 g ± 2 g) and Sprague-Dawley (SD) rats (120 g ± 12 g) were provided by the Animal Experimental Center of Southwestern Medical University (Animal License SCXK 2018-17), and all studies were approved by the Ethics Committee of Southwest Medical University. The precursor 2-(2-(vinyl sulfonyl)ethoxy)ethyl 4-nitrobenzene sulfonate was synthesized and characterized using the method described in the literature (24).

Synthesis of Intermediate Synthon [^{18}F]-((2-(2-Fluoroethoxy)Ethyl)Sulfonyl)Ethene

The [^{18}F]F$^-$ produced from cyclotron was trapped on a QMA cartridge and then eluted by a tetrabutyl ammonium bicarbonate (TBAB) solution in water and acetonitrile. The resulting tetrabutylammonium fluoride ([^{18}F]TBAF) solution was thoroughly dried by heating with anhydrous acetonitrile with N_2 blow three times and then redissolved in anhydrous acetonitrile. [^{18}F]TBAF (150 mCi in acetonitrile) was added to the solution of 2-(2-(vinyl sulfonyl)ethoxy)ethyl 4-nitrobenzene sulfonate (5 mg) in anhydride acetonitrile (80 μl) in a cap-sealed v-vial. The reaction mixture was heated at 85°C for 15 min. After cooling down, 1.0 ml of water was added, and the reaction mixture was loaded on HPLC for purification (column: type, AQ 5 μm; size, 4.6 mm × 250 mm, col. no. A6AD 10292; solvent A: 0.1% trifluoroacetic acid water; solvent B: 0.1% trifluoroacetic acid acetonitrile; 0–2 min: isocratic elution of 15% solvent B; 2–22 min, 15–95% of solvent B; flow rate: 3 ml/min). The desired product ^{18}F-VS has a retention time of 11 min, and the average radiochemical yield is 31%. The acetonitrile in the product was removed under vacuum, and the final product was reformulated with 1× phosphate-buffered saline (PBS, pH 7.3). The radiochemical purity analysis of ^{18}F-VS was performed using HPLC (**Supplementary Figure 2**).

RBCs Preparation

The rats were anesthetized with isoflurane, and 4 ml of blood was collected from the heart using an injection needle. Heparin (1,000 IU/kg body weight) was used to prevent blood clotting. After centrifugation (400 g for 10 min at 20°C), RBCs were located at the bottom of the tube. The plasma was separated from RBCs and stored for subsequent post-labeling stability studies. The buffy coating, which contains most of white blood cells and platelets, was removed, leaving the RBC layer undisturbed.

Optimization of Labeling Conditions

The separated RBCs and ^{18}F-VS solution (74 MBq, 600 μl) were mixed thoroughly and then divided into eight tubes and incubated separately at 0°C and 37°C ($n = 4$ in a group). The incubation time varied from 0 to 60 min, with 10-min intervals. At the corresponding incubation timepoint, 20 μl of the corresponding suspension was removed and mixed with an additional 100 μl PBS solution. Then, the mixture was subjected to centrifugation. Radioactivity of the supernatant and that of the erythrocyte sediment were measured separately. Optimal

conditions for labeling RBCs with ^{18}F-VS were determined and used thereafter for further evaluation.

Post-Labeling Stability of ^{18}F-VS-RBCs

Post-labeling stability was evaluated by calculating the radioactivity released from ^{18}F-VS-RBCs. The optimized labeling condition was detailed below: the centrifugal-washed RBC suspension (3 ml) and ^{18}F-VS solution (74 MBq in 600 μl) were mixed thoroughly and incubated at 37°C for 30 min. Then ^{18}F-VS-RBCs were washed three times with PBS (8 ml), and ^{18}F-VS-RBCs was isolated. To the purified ^{18}F-VS-RBCs, plasma solution was added and mixed. ^{18}F-VS-RBC suspension was divided into two tubes and incubated for 0 to 180 min (30-min intervals) at 0° and 37°C, respectively. After the corresponding incubation time, ^{18}F-VS-RBC suspension was cooled and uniformly resuspended. Then, the radioactivity of ^{18}F-VS-RBC suspension (10 μl) was measured ($n = 4$). The remaining ^{18}F-VS-RBC suspension was centrifuged at 450 g for 2 min, and the supernatant (10 μl) was sampled ($n = 4$). Radioactivity counts of the RBC suspension and supernatant were simultaneously measured. The release fraction was calculated according to the following formula: release fraction (%) = (radioactivity of the supernatant/radioactivity of initial RBC pellet) × 100%. The supernatant at 30 and 120 min timepoints were also analyzed by radio-HPLC. For the sample at the 120 min timepoint, elution from HPLC was collected per minute and counted by a gamma counter due to the low radioactivity.

Incubation of ^{19}F-VS With RBCs

^{19}F-VS was prepared with the previously reported method (26), which was then added to 5 μl RBCs in 2 ml saline to form a final concentration of ^{19}F-VS at 10 and 100 μM. No ^{19}F-VS was added in the blank control. All RBCs in the Petri dishes were incubated at 4°C, and the shape of RBCs was observed at the 2, 6, 12, and 24 h timepoints.

PET Imaging of Normal Rats

Normal SD rats (120 ± 12 g) were used in this study. All rats were anesthetized by inhalation of isoflurane, which was maintained throughout the imaging procedure. The imaging study was performed using a small-animal PET system. SD rats were placed on a fixed plate in the supine position for scanning. ^{18}F-VS-RBCs (24.8 ± 2.6 MBq and 400 μl) was injected through the tail vein. At the same time, images were continuously acquired for 60 min using the list mode. The list mode data were reconstructed using a dynamic sequence (30 frames, 60 s). After reconstruction, regions of interest (ROIs, mm^3) of the heart, blood vessels, and spleen were obtained using the software provided by the supplier (Inveon Research Workplace 4.2, Siemens). The values were presented as dose per gram of organ (% ID/g).

Imaging Study of the Rat Intra-Abdominal Hemorrhage Model

A glycerin enema was injected into the colon through the anus of rats to promote defecation about 30 min before the image acquisition. Then, 22.5 MBq of tracer (^{18}F-VS-RBCs with 1,000 IU/kg body weight heparin for anticoagulation) was injected through the tail vein and the PET/CT dynamic acquisition was started simultaneously. List mode data were acquired for 60 min. Under steady-state BPI conditions (10–15 min after injection), a 12-gauge lumbar puncture needle was used to manually puncture the colon wall through the anus to cause abdominal bleeding.

Image Analysis

For the abdominal hemorrhage model, the data acquisition and reconstruction were performed using the procedures described above. The ROI was drawn in the corresponding bleeding area. Radioactivity was presented using %ID/g, which was then then used to obtain the corresponding time–activity curve. The post-bleeding image (58–60 min after injection) and pre-bleeding image (10–20 min after injection) were subtracted with the PMOD software (Zurich, Switzerland) to measure the radioactivity of the bleeding site in the abdominal hemorrhage model. Negative values of the pixels in the subtracted image were replaced with zero values. Then, the total radioactivity of the abdominal hemorrhage image was presented as the percentage of the injected dose after excluding bladder radioactivity.

Statistical Analysis

Quantitative data were presented as mean ± standard deviation. Statistical analyses were performed using the SPSS Statistics 20.0 software package (IBM, Chicago, IL). The significance level was set to 0.05.

RESULTS AND DISCUSSION

^{18}F-VS Preparation

Similar to previous reports, the ^{18}F-VS was obtained through a nucleophile substitution of VS-ONs with [^{18}F]TBAF. The reaction mixture was purified using HPLC (**Supplementary Figure 1**), and the resulting ^{18}F-VS was obtained in 25–40% yield with ≥99% radiochemical purity and a retention time of 12.2 min (**Supplementary Figure 2**).

Evaluation Labeling Efficiency of ^{18}F-VS-RBCs

As a new BPI agent, it is important to optimize the labeling conditions to maximize the yield of ^{18}F-VS-RBCs. It is also important to understand the release profile from RBCs under different conditions (**Figure 1**).

The labeling efficiency (LE) of ^{18}F-VS on RBCs was higher with an incubation temperature of 37°C compared with 0°C. Prolonging the incubation time would increase the LE steadily, which would then reach a plateau. At 37°C, the LE is 76.93% (±1.66%), 80.10% (±1.08%), and 79.11% (±1.63%) at 10, 20, and 30 min, respectively. At 0°C, the yield is 58.17% (±2.18%), 63.12% (±5.67%), and 64.23% (±1.70%) at 10, 20, and 30 min, respectively. Further increasing the incubation time led to a decreased non-decay-corrected yield. Considering the increased radioactivity decay over time, we concluded that

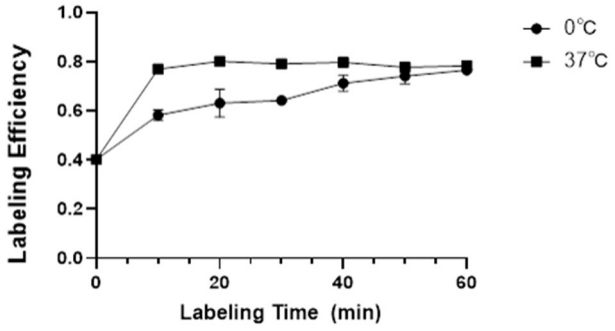


FIGURE 1 | The LE and its influencing factors. Relationship between labeling temperature, incubation time, and the LE of ^{18}F-VS-RBCs ($n = 4$).

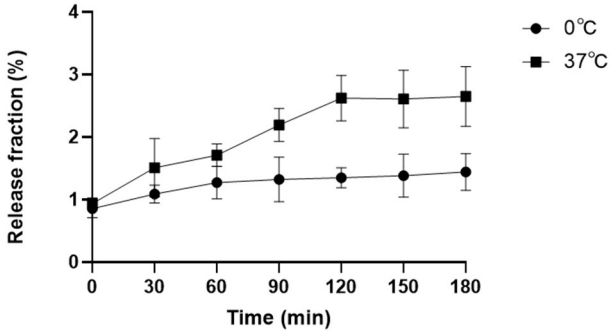


FIGURE 2 | Post-labeling stability analysis. The released radioactive fraction of ^{18}F-VS-RBCs when incubated in plasma at 37° or 0°C for 0–180 min ($n = 4$).

the optimal labeling condition for ^{18}F-VS-RBCs is 20 min incubation at 37°C.

Evaluation *in vitro* Stability of ^{18}F-VS-RBCs

To evaluate the *in vitro* stability of ^{18}F-VS-RBCs, a post-labeling stability experiment involving different storage temperatures was performed (**Figure 2**). In this study, the released ^{18}F-containing fraction from ^{18}F-VS-RBCs was evaluated for 180 min. Samples were incubated at 37° and 0°C, respectively, to simulate the temperature of the human body and the potential transportation/storage conditions using ice packs. ^{18}F-VS-RBC suspension (10 μl) was then taken for analysis at different incubation timepoints.

The release of ^{18}F was relatively small but noticeable. Incubation at 37°C for 0, 60, 120, and 180 min resulted in release fractions of 0.94% (±0.09%), 1.71% (±0.18%), 2.62% (±0.36%), and 2.65% (±0.48%), respectively. The corresponding samples incubated at 0°C had release fractions of 0.86% (±0.15%), 1.27% (±0.26%), 1.35% (±0.16%), and 1.44% (±0.29%), respectively. ^{18}F-VS-RBCs showed better stability when stored at 0°C compared with 37°C. This indicated that ^{18}F-VS-RBCs should be stored at 0°C between preparation and injection. The agent could be rewarmed before injection for *in vivo* research.

Overall, the released fractions of ^{18}F-VS-RBCs in the *in vitro* study were rather low. Thus, we concluded that ^{18}F-VS-RBCs were relatively stable *in vitro*. Moreover, HPLC analysis of the supernatant indicated that the released fraction only had a small amount of ^{18}F-VS in addition to some unknown radioactive fractions (**Supplementary Figure 3**). This observation indicated that ^{18}F-VS likely reacted with RBC through a covalent bond instead of passive absorption. As the released radioactive fraction may be excreted in the urine or absorbed by extravascular tissue after injection, imaging should be performed at an early timepoint if possible.

Evaluate the Toxicity of ^{19}F-VS on RBCs

To evaluate the toxicity of the tracer, the ^{19}F-VS was prepared and incubated with RBCs at 4°C in saline. The shape of RBCs was observed, and images of RBCs were recorded at each timepoint (**Supplementary Figure 5**). As shown in **Supplementary Figure 5**, the shape of RBCs stayed complete, and there was no obvious broken RBCs observed, indicating that ^{18}F-VS has no apparent toxicity on RBCs.

Evaluation of ^{18}F-VS-RBCs for PET Imaging

Under optimal conditions, the average time from drawing blood to intravenously injecting ^{18}F-VS-RBCs was ∼60 min. Microscopic examination showed that the morphology of ^{18}F-VS-labeled RBCs was normal without abnormal aggregation. The final ^{18}F-VS-RBC suspension (400 μl) had a radioactivity of 25.9 MBq (±3.7 MBq) and an LE of 70.09% (±0.61%) ($n = 4$).

PET Imaging of Normal Rats

In order to evaluate the distribution of ^{18}F-VS-RBCs in the cardiovascular system and the changes of radioactivity in the blood pool, we performed ^{18}F-VS-RBC imaging in normal rats. The maximum-intensity projection image and biodistribution of ^{18}F-VS-RBCs within 60 min after injection are shown in **Figure 3** and **Table 1**, respectively. The cardiovascular system in normal rats had a strong uptake of ^{18}F-VS-RBCs. A high uptake was obtained within 20 min, which was maintained stably at late time-points. Compared with that at 10 min, cardiac radioactivity at 60 min only decreased by ∼0.7%. The radioactivity of most organs remained relatively constant within 60 min. The results indicated that ^{18}F-VS-RBCs had a good stability *in vivo*. Within 60 min after tracer injection, atrium and spleen radioactivity was higher than that in the liver and lung. Urine excretion was observed. The PET imaging of the cardiovascular system in rats was clearly visualized with a low background ratio using ^{18}F-VS-RBCs. As shown in **Figure 3**, ^{18}F-VS-RBCs mainly stayed in the blood pool, suggesting good stability *in vivo*. At a late timepoint, urine activity was observed. As shown in **Supplementary Figure 3**, a small percentage of activity was released to the supernatant, which contains a hydrophilic motif. This may lead to the observed urine activity. Nonetheless, additional characterization would be done in a future study to further confirm it. Overall, ^{18}F-VS-RBCs hold a great potential to imaging the cardiovascular system considering its slow clearance and stability profile.

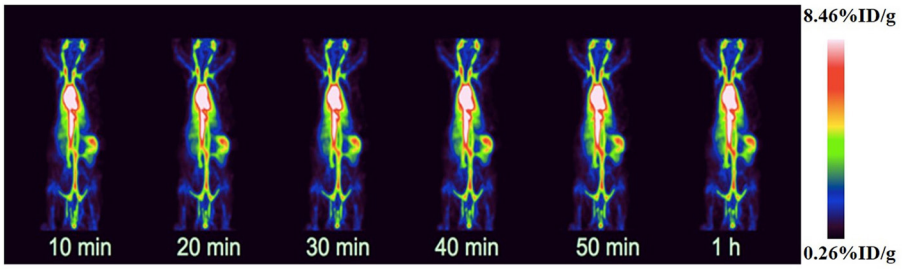


FIGURE 3 | Dynamic scan of normal rats using micro-PET/CT after injection with ^{18}F-VS-RBCs.

TABLE 1 | Biodistribution of ^{18}F-VS-RBCs at 60 min after injection in normal rats (n = 5/group).

Organ/tissue	Percentage of injected dose per gram of organ (%ID/g)					
	10 min	20 min	30 min	40 min	50 min	60 min
Atrium	7.79 ± 0.17	8.28 ± 0.18	8.16 ± 0.22	7.94 ± 0.18	7.91 ± 0.17	7.72 ± 0.16
Spleen	3.05 ± 0.11	2.83 ± 0.16	2.98 ± 0.16	2.99 ± 0.17	2.87 ± 0.15	2.82 ± 0.11
Lung	1.03 ± 0.07	1.09 ± 0.08	1.18 ± 0.09	0.98 ± 0.09	0.95 ± 0.07	0.93 ± 0.06
Liver	1.75 ± 0.03	1.63 ± 0.03	1.53 ± 0.04	1.50 ± 0.04	1.39 ± 0.02	1.36 ± 0.02
Kidney	1.62 ± 0.04	1.50 ± 0.05	1.43 ± 0.02	1.37 ± 0.02	1.38 ± 0.02	1.29 ± 0.03
Bladder	0.92 ± 0.14	2.96 ± 0.16	3.17 ± 0.18	4.91 ± 0.19	5.02 ± 0.17	7.37 ± 0.17

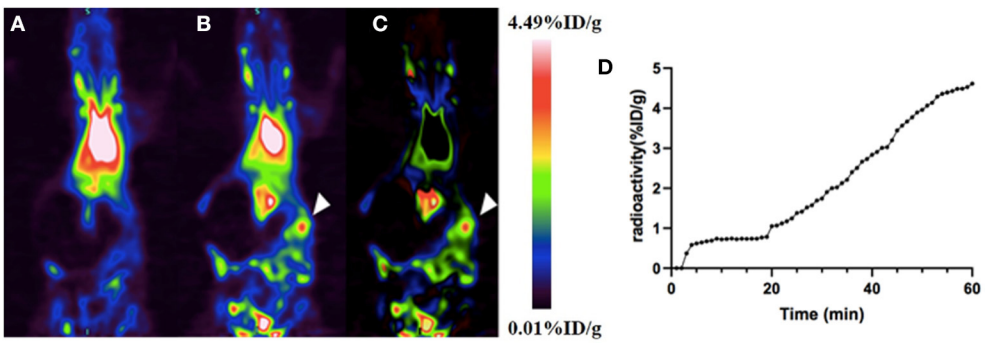


FIGURE 4 | The application of ^{18}F-VS-RBCs for PET imaging of the intra-abdominal hemorrhage model. The maximum-density projection image before **(A)** and after **(B)** bleeding and the subtraction image between the two **(C)**. The subtracted image showed heavy bleeding in the right abdomen (arrowheads). The time–activity curve of extravascular radioactivity in the abdominal hemorrhage area. Radioactivity continued to increase after manual puncture of the colon wall (18 min) **(D)**.

Imaging Study of the Intra-Abdominal Hemorrhage Model on Rats

A dynamic PET scan was performed on rats to evaluate the feasibility of using this BPI agent for the diagnosis of intra-abdominal hemorrhage. Dynamic PET imaging of the rat gastrointestinal bleeding model showed that shortly after the colon wall puncture, a high aggregation site appeared in the abdomen, indicating the tracer had extravasated due to bleeding (**Figures 4A–C**). Time–activity curves showed that the agent increased steadily at the bleeding site: radioactivity was stable within 10 min and continued to increase thereafter (**Figure 4D**). ^{18}F-VS-labeled RBCs successfully found the location and direction of abdominal bleeding over time. Although this study did not quantify the amount of bleeding, it can provide a rough estimation based on the blood radioactivity curve. Clearly, PET imaging of the rat intra-abdominal hemorrhage model demonstrated that ^{18}F-VS-RBCs hold a great potential for applications in gastrointestinal bleeding. Furthermore, without dietary restrictions, the labeling procedure of ^{18}F-VS-RBCs is simpler and shorter than the reported ^{18}F-FDG-RBCs. We would also like to point out that the potential limitations of ^{18}F-VS-RBCs include the requirement of clean space due to blood collection and *in vitro* labeling procedures. The radiation exposure toward operators could be high. Further improvements may focus on simplifying the labeling step by reducing the number of washing and streamlining the production process. Nonetheless, ^{18}F-VS-RBC PET of normal animal and intra-abdominal hemorrhage model suggested that the agent could represent a new BPI agent.

CONCLUSION

In this study, ^{18}F-VS was synthesized and successfully labeled RBCs for BPI. The resulting ^{18}F-VS-RBCs clearly visualized the cardiovascular system and extravascular blood in the abdominal hemorrhage model. Compared with existing cardiac blood pool agents, ^{18}F-VS-RBCs could be prepared using a readily available precursor and simple procedure. The agent has prolonged retention time in cardiac blood pool with a high-quality image. The agent's uptake was not affected by blood sugar, which eliminated the need of fasting. Because ^{18}F-VS-RBCs cannot penetrate the blood–brain barrier, it would be interesting to test its application in detecting bleeding in the brain in future studies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The animal study was reviewed and approved by The Affiliated Hospital of Southwest Medical University.

AUTHOR CONTRIBUTIONS

XZ and LW contributed to the study design, the labeling process, the imaging scan, the data analysis and they contributed equally to this paper. XZ wrote the manuscript and LW revised the manuscript. WF, YF, CZ, TX, and JC were responsible for the integrity of the data and the accuracy of the data analysis. TZ and LZ contributed to precursor synthesis, the labeling and evaluation of the tracer stability during the revision of the work. The YC and ZL was responsible for revising for important intellectual content. All authors have read and approved the final manuscript.

FUNDING

This study was funded by the Nuclear Medicine Innovation Transformation Platform (2019JDPT0004) and the Science and Technology Project of Sichuan Province (2018JY0230).

ACKNOWLEDGMENTS

The authors are grateful to the members of the Molecular Imaging Laboratory of the Affiliated Hospital of Southwest Medical University for their technical guidance, cooperation, and assistance with completing this research project. We would like to thank Mr. Lan Liu and Mr. Qiang Wan (Department of Nuclear Medicine, The Affiliated Hospital of Southwest Medical University) for assistance with the cyclotron operation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.646862/full#supplementary-material

REFERENCES

1. Werner DJ, Manner H, Nguyen-Tat M, Kloeckner R, Kiesslich R, Abusalim N, et al. Endoscopic and angiographic management of lower gastrointestinal bleeding: review of the published literature. *United Eur Gastroenterol J*. (2018) 6:337–42. doi: 10.1177/2050640617746299
2. Cheng HC, Khan MA, Bogdanov A Jr, Kwong K, Weissleder R. Relative blood volume measurements by magnetic resonance imaging facilitate detection of testicular torsion. *Invest Radiol*. (1997) 32:763–9. doi: 10.1097/00004424-199712000-00007
3. Massie BM, Kramer BL, Gertz EW, Henderson SG. Radionuclide measurement of left ventricular volume: comparison of geometric and counts-based methods, *Circulation*. (1982) 65:725–30. doi: 10.1161/01.CIR.65.4.725
4. Friedman BJ, Drinkovic N, Miles H, Shih WJ, Mazzoleni A, DeMaria AN. Assessment of left ventricular diastolic function: comparison of Doppler echocardiography and gated blood pool scintigraphy. *J Am Coll Cardiol*. (1986) 8:1348–54. doi: 10.1016/S0735-1097(86)80307-2
5. Ishikawa K, Saitoh M, Chida S. Detection of bladder hemangioma in a child by blood-pool scintigraphy. *Pediatr Radiol*. (2003) 33:433–5. doi: 10.1007/s00247-003-0895-8
6. de Eulate RG, Goñi I, Galiano A, Vidorreta M, Recio M, Riverol M, et al. Reduced cerebral blood flow in mild cognitive impairment assessed using phase-contrast MRI. *J Alzheimers Dis*. (2017) 58:585–95. doi: 10.3233/JAD-161222
7. Snelling LK, Helfaer MA, Traystman RJ, Rogers MC. Comparison of cerebral blood flow by radionuclide cerebral angiography and by microspheres in cats. *Crit Care Med*. (1992) 20:395–401. doi: 10.1097/00003246-199203000-00017
8. Rini JN, Bhargava KK, Tronco GG, Singer C, Caprioli R, Marwin SE, et al. PET with FDG-labeled leukocytes versus scintigraphy with 111In-oxine-labeled leukocytes for detection of infection. *Radiology*. (2006) 238:978–87. doi: 10.1148/radiol.2382041993
9. Hou G, Li X, Hou B, Zhou W, Cheng W. Lymphangioma on 68Ga-NOTA-Evans Blue PET/MRI. *Clin Nucl Med*. (2018) 43:553–5. doi: 10.1097/RLU.0000000000002129
10. Gutfilen B, Pellini MP, de Roure e Neder J, de Amarante Júnior JL, Evangelista MG, Fernandes SR, et al. 99mTc labeling white blood cells with a simple technique: clinical application, *Ann Nucl Med*. (1994) 8:85–9. doi: 10.1007/BF03164991
11. Schmidt KG, Rasmussen JW, Grove O, Madsen PE, Bargum R, Andersen D. Scintigraphic localization of intermittent gastrointestinal haemorrhage with indium-111-labelled platelets. *Scand J Gastroenterol*. (1983) 18:443–8. doi: 10.3109/00365528309181621
12. Winzelberg GG, McKusick KA, Froelich JW, Callahan RJ, Strauss HW. Detection of gastrointestinal bleeding with 99mTc-labeled red blood cells. *Semin Nucl Med*. (1982) 12:139–46. doi: 10.1016/S0001-2998(82)80005-6
13. Liu N, Wan Q, Cheng Z, Chen Y. Radionuclide-labeled peptides for imaging and treatment of CXCR4- overexpressing malignant tumors. *Curr Top Med Chem*. (2019) 19:17–32. doi: 10.2174/1568026619666190201094952
14. Pressacco J, Papas K, Lambert J, Paul Finn J, Chauny JM, Desjardins A, et al. Magnetic resonance angiography imaging of pulmonary embolism using agents with blood pool properties as an alternative to computed tomography to avoid radiation exposure. *Eur J Radiol*. (2019) 113:165–73. doi: 10.1016/j.ejrad.2019.02.007

15. Ito H, Yamamoto S, Saito K, Ikeda K, Hisada K. Quantitative estimation of hemorrhage in chronic subdural hematoma using the 51Cr erythrocyte labeling method. *J Neurosurg*. (1987) 66:862–4. doi: 10.3171/jns.1987.66.6.0862
16. Winzelberg GG, Froelich JW, McKusick KA, Waltman AC, Greenfield AJ, Athanasoulis CA, et al. Radionuclide localization of lower gastrointestinal hemorrhage. *Radiology*. (1981) 139:465–9. doi: 10.1148/radiology.139.2.6971455
17. Drescher R, Gröber S, Seifert P, Freesmeyer M. Differentiation of residual splenic tissue from neuroendocrine tumor metastasis on PET/CT with heat-damaged, Ga-68-oxine-labeled red blood cells. *Jpn J Clin Oncol*. (2021) 51:160–1. doi: 10.1093/jjco/hyaa066
18. Hladik WB III, Nigg KK, Rhodes BA. Drug-induced changes in the biologic distribution of radiopharmaceuticals. *Semin Nucl Med*. (1982) 12:184–218. doi: 10.1016/S0001-2998(82)80009-3
19. Hamilton RG, Alderson PO. A comparative evaluation of techniques for rapid and efficient *in vivo* labeling of red cells with [99mTc] pertechnetate. *J Nucl Med*. (1977) 18:1010–3.
20. Phelps ME, Hoffman EJ, Coleman RE, Welch MJ, Raichle ME, Weiss ES, et al. Tomographic images of blood pool and perfusion in brain and heart. *J Nucl Med*. (1976) 17:603–12.
21. Hoop B, Hnatowich DJ, Brownell GL, Jones T, McKusick KA, Ojemann RG, et al. Techniques for positron scintigraphy of the brain. *J Nucl Med*. (1976) 17:473–9.
22. Socan A, Petrik M, Kolenc Peitl P, Krošelj M, Rangger C, Novy Z, et al. On-cartridge preparation and evaluation of 68Ga-, 89Zr- and 64Cu-precursors for cell radiolabelling. *Nucl Med Biol*. (2019) 71:23–31. doi: 10.1016/j.nucmedbio.2019.04.001
23. Matsusaka Y, Nakahara T, Takahashi K, Iwabuchi Y, Nishime C, Kajimura M, et al. 18F-FDG-labeled red blood cell PET for blood-pool imaging: preclinical evaluation in rats. *EJNMMI Res*. (2017) 7:19. doi: 10.1186/s13550-017-0266-3
24. Wu Z, Li L, Liu S, Yakushijin F, Yakushijin K, Horne D, et al. Facile preparation of a thiol-reactive (18)F-labeling agent and synthesis of (18)F-DEG-VS-NT for PET imaging of a neurotensin receptor-positive tumor. *J Nucl Med*. (2014) 55:1178–84. doi: 10.2967/jnumed.114.137489
25. Shen BQ, Xu K, Liu L, Raab H, Bhakta S, Kenrick M, et al. Conjugation site modulates the in vivo stability and therapeutic activity of antibody-drug conjugates. *Nat Biotechnol*. (2012) 30:184–9. doi: 10.1038/nbt.2108
26. Zhang T, Cai J, Wang H, Wang M, Yuan H, Wu Z, et al. RXH-reactive 18F-vinyl sulfones as versatile agents for PET probe construction. *Bioconjug Chem*. (2020) 31:2482–7. doi: 10.1021/acs.bioconjchem.0c00487

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 22 December 2021
doi: 10.3389/fmed.2021.792677

Development and Validation of a Nomogram Based on ^{18}F-FDG PET/CT Radiomics to Predict the Overall Survival in Adult Hemophagocytic Lymphohistiocytosis

Xu Yang[1], Jun Liu[1], Xia Lu[1], Ying Kan[1], Wei Wang[1], Shuxin Zhang[1], Lei Liu[2], Hui Zhang[3], Jixia Li[4,5] and Jigang Yang[1]**

[1] Department of Nuclear Medicine, Beijing Friendship Hospital, Capital Medical University, Beijing, China, [2] Sinounion Medical Technology (Beijing) Co., Ltd., Beijing, China, [3] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China, [4] Department of Laboratory Medicine, School of Medicine, Foshan University, Foshan, China, [5] Department of Molecular Medicine and Pathology, School of Medical Science, The University of Auckland, Auckland, New Zealand

OPEN ACCESS

Edited by:
Xiaoli Lan, Huazhong University of Science and Technology, China

Reviewed by:
Alberto Miceli, Università di Genova, Italy
Salvatore Annunziata, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Italy

***Correspondence:**
Jigang Yang
yangjigang@ccmu.edu.cn
Jixia Li
j.li@auckland.ac.nz

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: *11 October 2021*
Accepted: *22 November 2021*
Published: *22 December 2021*

Citation:
Yang X, Liu J, Lu X, Kan Y, Wang W, Zhang S, Liu L, Zhang H, Li J and Yang J (2021) Development and Validation of a Nomogram Based on ^{18}F-FDG PET/CT Radiomics to Predict the Overall Survival in Adult Hemophagocytic Lymphohistiocytosis. Front. Med. 8:792677. doi: 10.3389/fmed.2021.792677

Purpose: Hemophagocytic lymphohistiocytosis (HLH) is a rare and severe disease with a poor prognosis. We aimed to determine if ^{18}F-fluorodeoxyglucose (^{18}F-FDG) PET/CT-derived radiomic features alone or combination with clinical parameters could predict survival in adult HLH.

Methods: This study included 70 adults with HLH (training cohort, $n = 50$; validation cohort, $n = 20$) who underwent pretherapeutic ^{18}F-FDG PET/CT scans between August 2016 and June 2020. Radiomic features were extracted from the liver and spleen on CT and PET images. For evaluation of 6-month survival, the features exhibiting $p < 0.1$ in the univariate analysis between non-survivors and survivors were selected. The least absolute shrinkage and selection operator (LASSO) regression analysis was used to develop a radiomics score (Rad-score). A nomogram was built by the multivariate regression analysis to visualize the predictive model for 3-month, 6-month, and 1-year survival, while the performance and usefulness of the model were evaluated by calibration curves, the receiver operating characteristic (ROC) curves, and decision curves.

Results: The Rad-score was able to predict 6-month survival in adult HLH, with area under the ROC curves (AUCs) of 0.927 (95% CI: 0.878–0.974) and 0.869 (95% CI: 0.697–1.000) in the training and validation cohorts, respectively. The radiomics nomogram combining the Rad-score with the clinical parameters resulted in better performance for predicting 6-month survival than the clinical model or the Rad-score alone. Moreover, the nomogram displayed superior discrimination, calibration, and clinical usefulness in both the cohorts.

Conclusion: The newly developed Rad-score is a powerful predictor for overall survival (OS) in adults with HLH. The nomogram has great potential for predicting 3-month, 6-month, and 1-year survival, which may timely guide personalized treatments for adult HLH.

Keywords: hemophagocytic lymphohistiocytosis, radiomics, prognosis, ^{18}F-FDG PET/CT, nomogram

INTRODUCTION

Hemophagocytic lymphohistiocytosis (HLH) is a syndrome of severe immune activation and dysregulation characterized by hyperactive cytotoxic T lymphocytes, natural killer (NK) cells, and macrophages leading to cytokine storm and immune-mediated multiple organ failure (1, 2). Historically, HLH has been classified as primary or familial HLH driven by underlying genetic defects in cytotoxic immune function or as secondary or reactive HLH caused by infections [e.g., Epstein-Barr virus (EBV), cytomegalovirus (CMV), HIV, and coronavirus disease 2019 (COVID-19)], malignancies (e.g., hematologic malignancies), and autoimmune diseases (e.g., macrophage activation syndrome) (1). Emerging evidence demonstrated that HLH may occur in patients of any age and is most often driven by an integration of genetic defects and acquired exposures (3, 4). Primary HLH occurs in 1/50,000–1/100,000 live-born children, while secondary HLH occurs in older children and adults (1). The precise incidence of adult HLH is still unknown, but it accounts for ∼40% of all HLH (1, 5). The frequent manifestations are intermittent fever, hepatosplenomegaly, lymphadenopathy, liver injury, cytopenia, hypertriglyceridemia, hyperferritinemia, and hemophagocytosis (1). Because of few data and/or no prospective studies for adult HLH, pediatric data are often generalized to guide diagnostic, therapeutic, and prognostic decision-making in adults (1, 6). In general, adults have poorer outcome than children even with aggressive therapy, with a median survival of 4 months (1). The principal reasons for mortality are multiorgan failure, hemorrhage, and sepsis, which can be treated properly if diagnosed early (7). Therefore, identifying poor prognosis in adult HLH is crucial for risk stratification and therapeutic decision-making. Recently, it has been reported that clinical and laboratory markers are correlated with survival in adult HLH including age, platelet, fibrinogen, albumin, serum ferritin, alanine aminotransferase (ALT), and malignancy (1, 2, 8, 9). But none of them can be a single effective prognostic factor as a result of poor sensitivity and/or specificity. Hence, it would be of great utility to build a predictive model to precisely evaluate the prognosis in adult HLH based on multiple indictors.

^{18}F-fluorodeoxyglucose (^{18}F-FDG) PET/CT has been employed for detecting underlying malignancy and predicting prognosis of adult HLH (10–12). One of the most common PET/CT finding is hepatosplenomegaly with diffusely increased FDG uptake, which contains a great deal of information reflecting disease status in adult HLH (13, 14). Radiomics can convert medical images into quantitative data and subsequently analyze these data for prognosis prediction by high-throughput computing. PET/CT radiomic features have been explored to predict outcome in malignancies such as lymphoma and lung cancer (15–17). It has been suggested that the quantitative PET parameters of spleen are independent prognostic factors (11, 12), but whether PET/CT radiomic features extracted from liver and spleen can be applied for outcome prognostication in adult HLH is unclear yet. Therefore, the first aim of this study was to establish a PET/CT radiomics score (Rad-score) for predicting 6-month survival in adult HLH and the second aim was to combine the Rad-score with clinical parameters, in order to develop a nomogram for predicting individual prognosis accurately and reliably.

MATERIALS AND METHODS

Patients

This retrospective study was approved by Institutional Review Board of Beijing Friendship Hospital of Capital Medical University and the requirement of a written informed consent was waived. The medical records of 185 consecutive adult patients (age $\geq$ 18 years) with a diagnosis of HLH were reviewed from August 2016 to June 2020. The diagnostic criteria of HLH were in accordance with HLH-2004 protocol, which requires five of the following eight criteria: (1) fever; (2) splenomegaly; (3) cytopenia affecting $\geq$ 2 lineages (Hemoglobin (HGB) $<$ 9 g/dl, platelets $< 100 \times 10^9$/L, neutrophils $< 1.0 \times 10^9$/L); (4) serum triglyceride $\geq$ 265 mg/dl and/or fibrinogen $\leq$ 150 mg/dl; (5) hemophagocytosis in bone marrow, spleen, lymph nodes, or liver; (6) low or absent NK cell activity; (7) Ferritin $\geq$ 500 μg/l; and (8) soluble interleukin-2 receptor (soluble CD25) $\geq$ 2,400 U/ml (18). The exclusion criteria included: patients with receiving chemotherapy before ^{18}F-FDG PET/CT scan ($n = 114$) or incomplete follow-up ($n = 1$). Consequently, a total of 70 patients were included in this study. All the patients received personalized treatments in the Department of Hematology and were followed-up for at least 180 days with a median of 353 days. These cases were randomly divided into the training ($n = 50$) and validation cohorts ($n = 20$) with a ratio of 5:2.

Clinical Data Collection

Clinical parameters including age, gender, malignancy, EBV infection, hemophagocytosis, and laboratory variables [white blood cell, absolute neutrophil, hemoglobin, platelet, C-reactive protein (CRP), ALT, aspartate aminotransferase (AST), triglycerides, serum ferritin, fibrinogen, erythrocyte sedimentation rate, and lactate dehydrogenase] were obtained from medical records (**Table 1**). All the laboratory and radiological data were collected before initial HLH-specific therapy. The most likely trigger of secondary HLH (malignancy, infection, autoimmune, and idiopathic) was determined by assessment and medical evidence of physician.

^{18}F-Fluorodeoxyglucose PET/CT Imaging Acquisition, Segmentation, and Feature Extraction

^{18}F-fluorodeoxyglucose PET/CT was performed on a Siemens biography mCT PET/CT scanner (Siemens Healthineers, Erlangen, Germany). Patients were instructed to fast for at least 6 h, accompanied by blood glucose <11.1 mmol/l. Then, ^{18}F-FDG (4.4 MBq/kg) was injected intravenously. After a 60-min uptake time, low-dose CT scan was executed for visualization of anatomic structures and attenuation correction, with 140 keV, automatic mAs, and a slice thickness of 3 mm. The whole-body PET scan was carried out with 2.5 min per bed position using three-dimensional (3D) mode immediately after a whole-body CT scan. Images were reconstructed with an iterative reconstruction algorithm.

TABLE 1 | Clinical characteristics of patients in the training and validation cohorts.

Variables	Overall ($n = 70$)	Training cohort ($n = 50$)	Validation cohort ($n = 20$)	p value
Gender				
Male	36 (51.4%)	23 (46.0%)	13 (65.0%)	0.151
Female	34 (48.6%)	27 (54.0%)	7 (35.0%)	
Age (years)	38 (27–55)	36(24–53)	44(32–60)	0.194
Malignancy				
Yes	22 (31.4%)	15 (30.0%)	7 (35.0%)	0.684
No	48 (68.6%)	35 (70.0%)	13 (65.0%)	
T cell neoplasms				
Yes	11 (15.7%)	8 (16.0%)	3 (15.0%)	0.917
No	59 (84.3%)	42 (84.0%)	17 (85.0%)	
EBV infection				
Positive	34 (48.6%)	23 (46.0%)	11 (55.0%)	0.496
Negative	36 (51.4%)	27 (54.0%)	9 (45.0%)	
Hemophagocytosis				
Yes	48 (68.6%)	34 (68.0%)	14 (70.0%)	0.871
No	22 (31.4%)	16 (32.0%)	6 (30.0%)	
WBC ($\times 10^9$/L)	4.11 (1.63–6.9)	3.92(1.63–6.02)	5.47 (1.91–10.16)	0.101
ANC ($\times 10^9$/L)	2.17 (0.94–4.41)	2.13 (0.98–3.82)	2.99 (1.10–6.77)	0.108
HGB (g/L)	92.5 (73.3–108.5)	93 (80–105)	90 (72–111)	0.691
PLT ($\times 10^9$/L)	85.5 (55.0–179.5)	86 (55–150)	137 (56–194)	0.179
CRP (mg/L)	26 (5–56)	26 (5–49)	34 (11–53)	0.824
ALT (U/L)	53 (32–102)	45 (25–102)	69 (49–96)	0.487
AST (U/L)	68 (33–122)	63 (33–124)	70 (45-118)	0.460
TG (mmol/L)	1.97 (1.43–2.61)	1.79 (1.38–2.47)	2.25 (1.89–3.00)	0.107
SF (ng/ml)	1714.0 (625.6–4075.0)	1450.5 (773.1–3428.5)	3356.5 (1031.3–5359.0)	0.915
FBG (g/L)	2.23 (1.43–3.24)	2.19 (1.42–3.20)	2.63 (1.70–3.13)	0.562
ESR (mm/h)	22 (10–44)	22 (10–38)	24 (12–49)	0.342
LDH (U/L)	547 (343–940)	616 (343–926)	499 (398–764)	0.606

Data are expressed as median (interquartile range) or number (the proportion of sample size).
EBV, Epstein-Barr virus; WBC, white blood cell; ANC, absolute neutrophil count; HGB, hemoglobin; PLT, platelet count; CRP, C-reactive protein; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TGs, triglycerides; SF, serum ferritin; FBG, fibrinogen; ESR, erythrocyte sedimentation rate; LDH, lactate dehydrogenase.

The entire liver and spleen on CT images were defined as the regions of interest (ROIs), which were delineated by two experienced nuclear radiologists with a validated semi-automatic approach using (3D Slicer^{TM} software, Boston, Massachusetts, United States) (version 4.10.0, http://www.slicer.org) (**Figure 1**) (19). Moreover, the ROIs were resampled exploiting B-spline interpolation in order for mapping those onto the PET images. In consequence, the ROIs had the matching pixel spacing with the PET images.

Radiomic Feature Extraction

Radiomic features were extracted from ^{18}F-FDG PET and CT images separately, using pyradiomics that is an open-source Python package (20). These included first-order features ($n = 18$), shape features ($n = 14$), gray level co-occurrence matrix (GLCM) features ($n = 24$), gray level run length matrix (GLRLM) features ($n = 16$), gray level size zone matrix (GLSZM) features ($n = 16$), neighboring gray tone difference matrix (NGTDM) features ($n = 5$), and gray level dependence matrix (GLDM) features ($n = 14$). Image processing utilized wavelet filtering, square, square root, logarithm, exponential, and gradient. Overall, a total of 5,264 radiomic features (4 $\times$ 1,317) were obtained from liver and spleen on PET and CT images.

Radiomic Feature Selection and the Rad-Score Construction

Our workflow is shown in **Figure 1**. Firstly, the univariate analysis (t-test for normally distributed variables or the Mann–Whitney U test for skewed distributed variables) was used to compare differences of radiomic features between non-survivors and survivors at 180 days in the training set. The total of 384 features with p-values < 0.1 were retained for further analysis. Next, the least absolute shrinkage and selection operator (LASSO) algorithm was applied to select the optimal features among 384 features in the training set, adding L1 regularization term to a least square algorithm for data dimension reduction. Because of imbalanced datasets, the synthetic minority oversampling technique (SMOTE) was used to improve random oversampling in the training set.

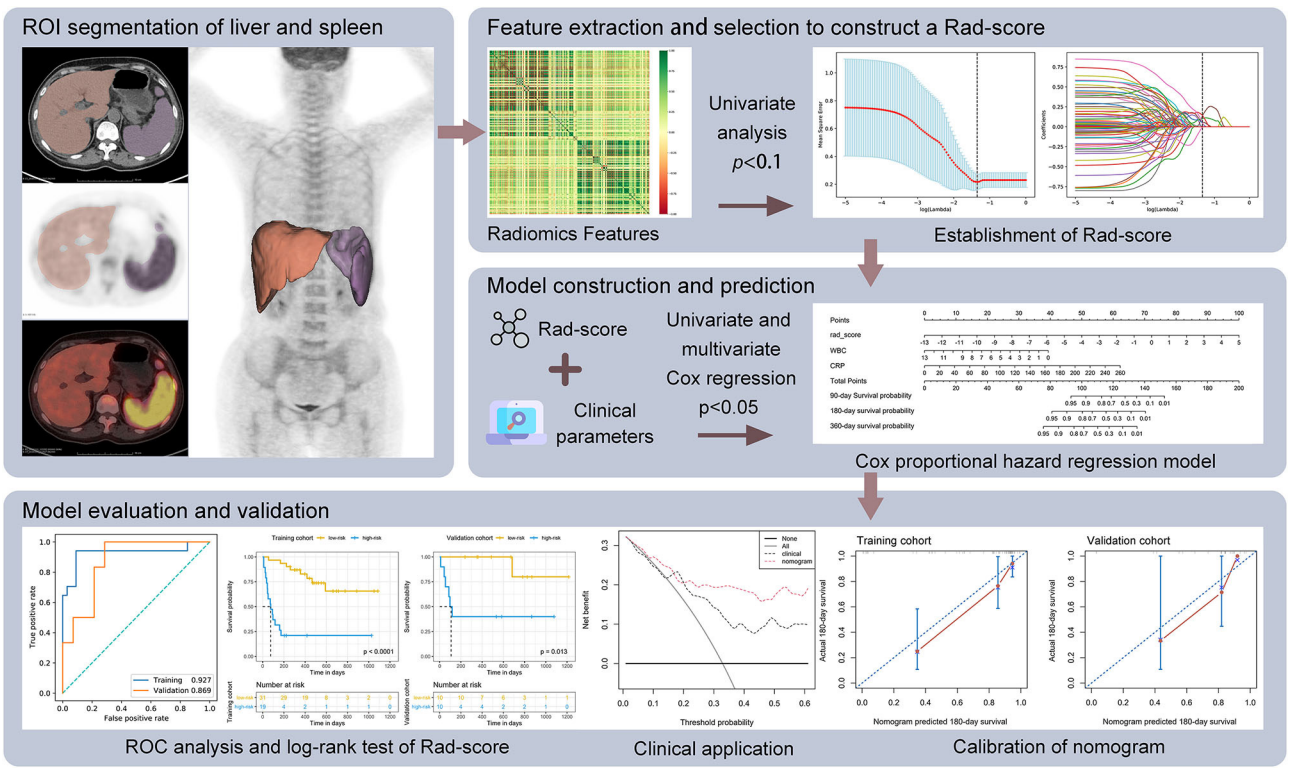


FIGURE 1 | Workflow of radiomics analysis.

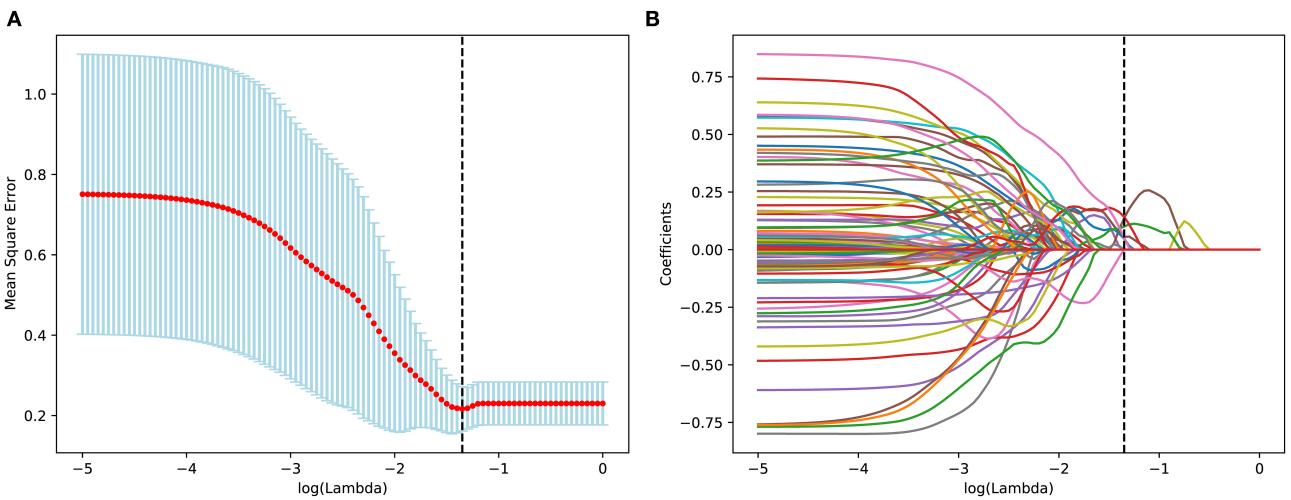


FIGURE 2 | Feature selection for the prediction using the least absolute shrinkage and selection operator (LASSO) regression model, tuning parameter (λ) selection in the LASSO model involved the use of tenfold cross-validation **(A)**. In the coefficient profiles of the radiomics features for OS prediction, a value of Lambda = 0.044668 was selected as the optimal value **(B)**.

An individualized Rad-score was calculated from a linear combination of the selected features weighted by their respective coefficients. The receiver operating characteristic (ROC) curve was employed to evaluate the prediction accuracy and determine the optimal threshold of the Rad-score. All the patients were divided into high- and low-risk groups according to the maximum Youden index of the ROC curve. The potential association of the Rad-score with overall survival (OS) was evaluated by the Kaplan–Meier survival analysis and the log-rank test in the training and validation cohorts.

TABLE 2 | Comparison of the radiomics features between 180-day survivors and non-survivors in the training cohort.

Radiomic features	Survivors ($n = 33$)	Non-survivors ($n = 17$)	p value
spleen_CT_wavelet-LHL_ngtdm_Contrast	0.0024 (0.0016–0.0045)	0.0037 (0.0020–0.0050)	0.072
spleen_CT_wavelet-HHH_glszm_Gray Level Non-Uniformity Normalized	0.4259 (0.3884–0.5000)	0.4102 (0.3333–0.4404)	0.045
spleen_PET_squareroot_firstorder_Kurtosis	4.8251 (3.9777–6.1500)	5.7335 (4.1330–7.8258)	0.035
spleen_PET_wavelet-LHL_glszm_Size Zone Non-Uniformity Normalized	0.1927 (0.1641–0.2355)	0.2388 (0.2000–0.2800)	0.004
liver_CT_wavelet-HHH_glcm_Imc2	0.0712 (0.0682–0.0725)	0.4500 (0.4222–0.5185)	0.082
liver_PET_wavelet-HHL_glszm_Small Area Emphasis	0.3890 (0.2778–0.5284)	0.3890 (0.2778–0.5284)	0.084

Data are expressed as median (interquartile range).

TABLE 3 | Comparison of the Rad-score between 180-day survivors and non-survivors in both the training and validation cohorts.

	Training cohort ($n = 50$)		p value
	Survivors ($n = 33$)	Non-survivors ($n = 17$)	
Rad-score	−1.0608 (−1.9723–−0.6384)	1.6386 (0.5678–2.9977)	<0.001
	Validation cohort ($n = 20$)		p value
	Survivors ($n = 14$)	Non-survivors ($n = 6$)	
Rad-score	−1.3595 (−2.1365–0.0290)	1.3763 (−0.0232–2.9420)	0.011

Data are expressed as median (interquartile range).

Clinical Variables Selection and Nomogram Creation

To build a powerful model and a robust nomogram for the survival prediction, the clinical prognostic factors were chosen by the univariate Cox regression analyses ($p < 0.05$). Then, the Rad-score and the strong clinical indicators were incorporated to establish the multivariate Cox regression model that was visualized by a nomogram. The Harrell's concordance-index (C-index) was employed to assess the model performance and calibration curves were plotted to enhance the predictive precision of nomogram. Similarly, a clinical model was established with clinical information alone by the multivariate Cox regression analysis. Three different types of predictive models (clinical variables, the Rad-score, and their combinations) were evaluated by the C-index. Decision curve analysis (DCA) was utilized to assess the clinical usefulness of the models.

Statistical Analyses

Continuous variables are presented as medians with interquartile ranges and categorical variables are presented as frequencies and percentages. OS was defined as the time from the initial diagnosis of HLH to the date of death from any cause or deadline of follow-up. All the p-values were two-sided, with a significant level of <0.05. Statistical analyses were performed with Python (version 3.7.8, www.python.org) and R (version 4.0.3, www.r-project.org). The Python packages "sklearn," "numpy," and "pandas" were used for the LASSO binary logistic regression and the ROC curve; the "scipy" was for analyzing statistical properties; and the "imblearn" was for analyzing SMOTE. The R package "rms" was employed to create nomograms.

RESULTS

Baseline Clinical Characteristics of Patients

A total of 70 adults with HLH were included in this study who fulfilled the inclusion criteria. There were 36 males and 34 females and the median age at diagnosis was 38 years (range: 18–79 years). The baseline characteristics of all the patients are shown in **Table 1**.

The possible triggers of HLH in these patients were as follows: 35 (50.0%) infections, 22 (31.4%) malignancies, 7 (10%) autoimmune diseases, and 6 (8.6%) unknown disorders. In the 35 cases with infectious disorders, viral infections were the most common cause with 23 (65.7%) EBV, 3 (8.6%) CMV, and 5 (14.3%) other viruses. Bacterial infections were identified in 4 (11.4%) patients. Among 22 malignancy-associated HLH cases, diffuse large B-cell lymphoma ($n = 6$, 27.3%) and NK/T-cell lymphoma ($n = 6$, 27.3%) were the most frequent triggers. The other malignancy-associated patients with HLH were two classical Hodgkin's lymphoma, two peripheral T-cell lymphoma, two unclassified T-cell lymphoma, one follicular lymphoma, one non-Hodgkin B cell lymphoma, one anaplastic large cell lymphoma, and one acute lymphocytic leukemia. Concomitant malignancies and EBV infection were found in 11 (15.7%) patients. Adult Still's disease ($n = 5$, 71.4%) was the most common diagnosis among autoimmune diseases and the other two patients were diagnosed with systemic lupus erythematosus and undifferentiated systemic rheumatic disease, respectively.

The baseline characteristics in the training and validation cohorts are also given in **Table 1**. Obviously, the clinical variables had no differences between the two cohorts ($p > 0.05$). After a

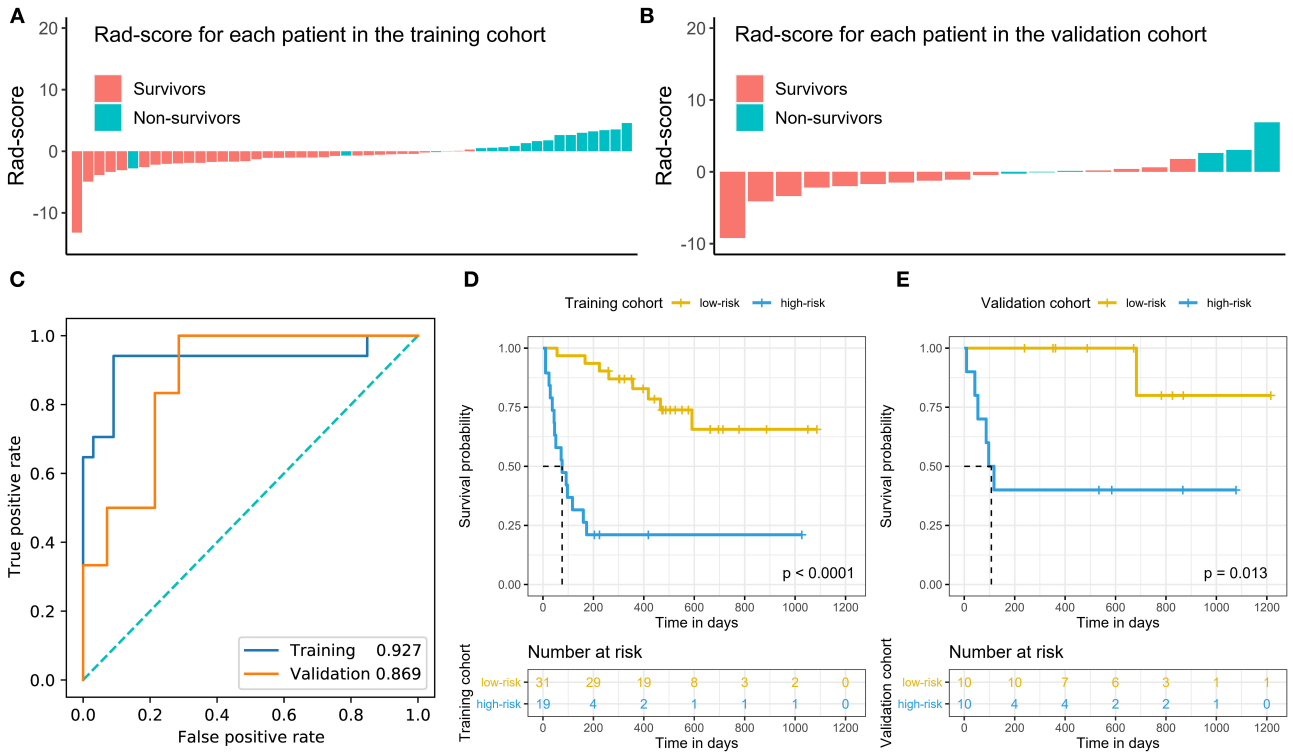


FIGURE 3 | Rad-score of patients in the training and validation cohorts **(A,B)** and time-dependent ROC analysis of Rad-score at 180 days **(C)**. Kaplan-Meier survival analysis with the best cutoff value of the Rad-score in the training cohort **(D)** and validation cohort **(E)**. We calculated *p* values using the log-rank test.

median follow-up of 353 days (range: 9–1,216 days), 30 patients (42.8%) had died.

Radiomic Feature Selection and the Rad-Score Construction and Evaluation

The optimal radiomic features were selected by the LASSO algorithm and 10-fold cross-validation (**Figure 2**). Eventually, six features were extracted to construct the Rad-score in the training set, which included 3 CT features and 3 PET features. The selected features were spleen_CT_wavelet-HHH_GLSZM_gray level non-uniformity normalized, spleen_CT_wavelet-LHL_NGTDM_contrast, liver_CT_wavelet-HHH_GLCM_informational measure of correlation (IMC) 2, spleen_PET_square root_first order_kurtosis, spleen_PET_wavelet-LHL_GLSZM_size zone non-uniformity normalized, and liver_PET_wavelet-HHL_GLSZM_small area emphasis. Among these features, there were four from spleen and two were from liver. The Rad-score for each patient was calculated by the following formula:

Rad-score = 13.762 + 300.60 × spleen_CT_wavelet-LHL_NGTDM_contrast − 9.1753 × spleen_CT_wavelet-HHH_GLSZM_gray level non-uniformity normalized − 0.17938 × spleen_PET_square root_first order_kurtosis + 13.305 × spleen_PET_wavelet-LHL_GLSZM_size zone non-uniformity normalized − 246.77 × liver_CT_wavelet-HHH_GLCM_IMC 2 + 4.2599 × liver_PET_wavelet-HHL_GLSZM_small area emphasis.

The median and the interquartile range for the selected radiomics features in the training cohort are shown in **Table 2**. The Rad-score in the training and validation cohorts are shown in **Table 3**. Not surprisingly, the Rad-score had notable difference between non-survivors and survivors in the training ($p < 0.001$) and validation cohorts ($p = 0.011$). Particularly, non-survivors had the higher Rad-score than survivors in the training (Rad-score = 1.6386 vs. −1.0608) and validation cohorts (Rad-score = 1.3763 vs. −1.3595). The Rad-score for individuals in the both the cohorts is shown in **Figures 3A,B**.

In addition, the Rad-score had good predictive power for survival forecast at 180 days and its area under the ROC curves (AUCs) in distinguish high-risk status were 0.927 (95% CI: 0.879–0.975) in the training set and 0.869 (95% CI: 0.684–1.000) in the validation set (**Figure 3C**). The best cutoff with maximum Youden index was −0.3 and, therefore, patients were divided into high- and low-risk groups according to the Rad-score in the both the cohorts. The Kaplan–Meier curves and the log-rank test found that patients in low-risk category had a better prognosis than those in high-risk category in the training and validation cohorts ($p < 0.05$) (**Figures 3D,E**).

Strong Predictor Selection and Model Establishment and Assessment

The univariate Cox regression analysis showed that 6 parameters were significantly associated with OS including the Rad-score, T-cell neoplasms, white blood cell, hemoglobin, platelet count,

TABLE 4 | The univariate and multivariate Cox hazards regression analysis of OS in the training cohort.

Variables	Univariate analysis		Multivariate analysis	
	Hazard ratio (95% CI)	*p* value	Hazard ratio (95% CI)	*p* value
Rad-score	1.683 (1.312–2.158)	<0.001	1.522 (1.221–1.896)	<0.001
Gender	1.930 (0.844–4.412)	0.119		
Age	1.017 (0.992–1.044)	0.185		
Malignancy	1.569 (0.676–3.640)	0.295		
T cell neoplasms	3.304 (1.292–8.448)	0.013		
EBV infection	2.240 (0.966–5.195)	0.060		
Hemophagocytosis	1.176 (0.483–2.861)	0.721		
WBC	0.790 (0.661–0.945)	0.010	0.796 (0.661–0.957)	0.015
ANC	0.825 (0.656–1.039)	0.102		
HGB	0.971 (0.951–0.992)	0.006		
PLT	0.991 (0.984–0.997)	0.007		
CRP	1.007 (1.002–1.013)	0.009	1.018 (1.010–1.027)	<0.001
ALT	0.996 (0.990–1.002)	0.230		
AST	0.999 (0.997–1.001)	0.518		
TG	0.827 (0.500–1.366)	0.458		
SF	1.000 (1.000–1.000)	0.202		
FBG	0.900 (0.662–1.224)	0.502		
ESR	0.997 (0.981–1.013)	0.702		
LDH	1.000 (0.999–1.001)	0.756		

OS, overall survival; EBV, Epstein-Barr virus; WBC, white blood cell; ANC, absolute neutrophil count; HGB, hemoglobin; PLT, platelet count; CRP, C-reactive protein; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TGs, triglycerides; SF, serum ferritin; FBG, fibrinogen; ESR, erythrocyte sedimentation rate; LDH, lactate dehydrogenase.

TABLE 5 | The multivariate Cox hazards regression analysis of OS in the training cohort without the Rad-score.

Variables	Hazard ratio (95% CI)	*p* value
T cell neoplasms	5.800 (1.957–17.187)	0.002
HGB	0.999 (0.946–0.986)	<0.001
PLT	0.994 (0.987–1.000)	0.050

HGB, hemoglobin; PLT, platelet count.

TABLE 6 | Model performance.

Model	Training cohort		Validation cohort	
	C-index	95% CI	C-index	95% CI
Rad-score	0.795	0.695–0.895	0.752	0.591–0.913
Clinical model	0.765	0.665–0.865	0.762	0.527–0.997
Combined radiomics model	0.831	0.749–0.913	0.810	0.657–0.963

C-index: Harrell's concordance-index.

and CRP ($p < 0.05$; **Table 4**). The multivariate analysis displayed that the Rad-score, white blood cell, and CRP were consistently strong predictors (**Table 4**), which were used to build the combined model. When the Rad-score was excluded, three variables (T-cell neoplasms, hemoglobin, and platelet count) were independent prognostic factors among clinical parameters (**Table 5**). Likewise, these three prognostic factors were used to build the clinical model.

To assess the performance of models in predicting prognosis, the C-indices of three types of models were shown in **Table 6**. The Rad-score model had acceptable predictive ability with C-indices of 0.795 (95% CI: 0.695–0.895) and 0.752 (95% CI: 0.591–0.913) in the training and validation cohorts, respectively. The C-indices of the clinical model were of 0.765 (95% CI: 0.665–0.865) and 0.762 (0.527–0.997) in the training and validation cohorts, respectively. It was noticeable that the combined radiomics model had the highest C-indices, with 0.831 (95% CI: 0.749–0.913) and 0.810 (95% CI: 0.657–0.963) in both the training and validation cohorts, sequentially. The curves of decision-curve analysis (DCAs) indicated that the combined radiomics model provided more net clinical benefit than clinical model with a threshold > 0.25 (**Figure 4**).

Personalized Nomogram Establishment and Validation

Given that the combined model possessed synergetic power for survival prediction, the personalized nomogram was constructed by incorporating all the three independent prognostic factors (the Rad-score, white blood cell, and CRP) (**Figure 5A**), which can visualize the prediction outcome and the proportion of each factor. The calibration curves demonstrated good agreements between the predicted and observed values in the training and validation cohorts, indicating that the nomogram was able to precisely predict 6-month survival (**Figures 5B,C**).

DISCUSSION

Timely diagnosis and prognosis are critical for HLH considering that the early and proper administration of an efficacious

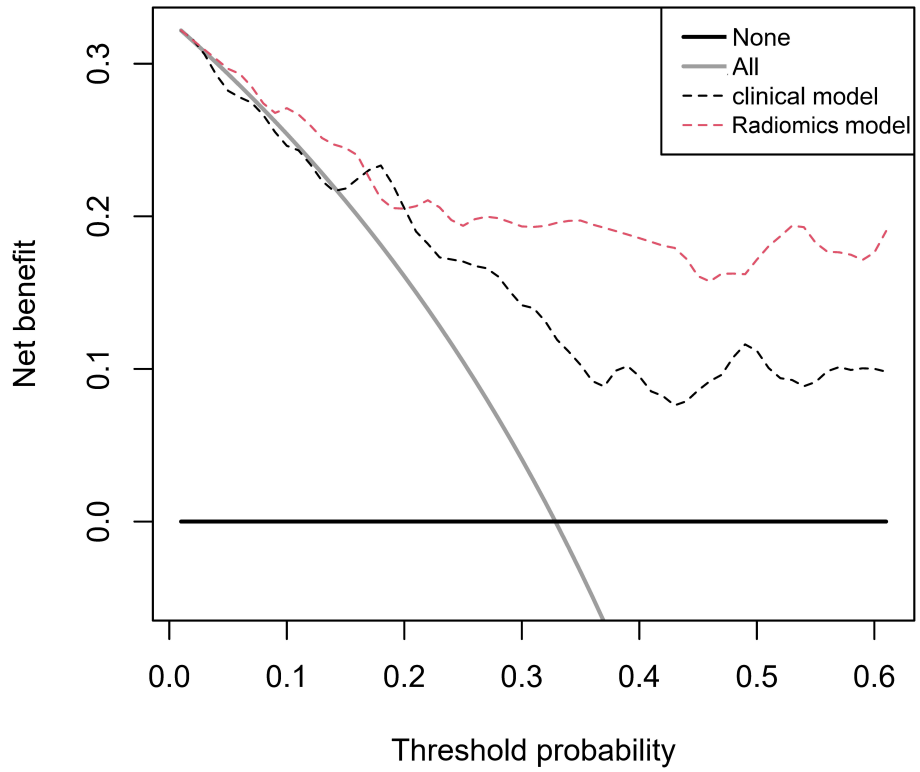


FIGURE 4 | Decision-curve analysis for the radiomics model and clinical model. The threshold probability represents the predicted 180-day risk of death for recommending aggressive treatment.

therapy can improve survival. In this study, the novel prognostic factors and predictive models associated with 6-month survival in adult patients with HLH are reported *via* pretherapeutic ^{18}F-FDG PET/CT radiomics analysis. The Rad-score and the combined prediction model (the Rad-score and clinical variable combination) have been developed for quantitative identification of the adults with HLH at high risk of death within 6 months in 70 patients.

^{18}F-fluorodeoxyglucose PET/CT is a whole-body scan containing both the metabolic and anatomical information, which has been recommended for identifying possible triggers and suitable biopsy sites in secondary HLH (6). However, ^{18}F-FDG PET/CT findings are non-specific, since inflammatory response and malignant lesions have the same manifestation that is hypermetabolism. In HLH, ^{18}F-FDG PET/CT often shows diffusely increased FDG uptake in spleen, liver, and bone marrow with or without focal lesions and hypermetabolic lymph nodes. Increasing evidence demonstrated that these non-specific presentations can be used to assess systemic inflammatory response and have potential for prognosis prediction in HLH (21). For instance, the FDG uptake of spleen and bone marrow has been considered as prognostic factors in adult patients with HLH (11, 12, 22). More importantly, spleen and liver, components of the reticuloendothelial system, are the most frequent abnormal signs in HLH (23). Our data proved that 6 radiomic features from spleen and liver were linked with the prognosis of adult HLH, thus utilized for establishment of the Rad-score. Among six radiomic features, two-thirds (4/6) were derived from spleen including GLSZM size zone non-uniformity normalized feature and kurtosis of spleen PET and GLSZM gray level non-uniformity normalized feature and NGTDM contrast of spleen CT. It is well-known that spleen is the largest secondary lymphoid organ and a site where immune responses can be controlled by activated immune cells. As splenomegaly is one of the diagnostic criteria of HLH, the hypermetabolic spleen has been discovered to be correlated with high inflammatory response and cytokine activity (24, 25). One recent report suggested that spleen FDG uptake may provide useful information for predicting in-hospital mortality in autoimmune diseases including HLH (26). Another study of 43 patients with secondary HLH found that the ratio of spleen to mediastinum in the average standardized uptake value (SUV) was an independent predictor for survival (12). In consistent with these statements, our findings indicated that radiomic features of spleen possessed a powerful predictive ability for 6-month survival in adult patients with HLH. The rest of two radiomic features were extracted from liver including GLCM IMC2 (Informational Measure of Correlation) of liver CT and GLSZM small area emphasis of liver PET. Radiomics have showed great value in characterization of diffuse liver diseases such as non-alcoholic steatohepatitis and chronic hepatitis B (27). In addition, the well-known liver enzymes, AST and

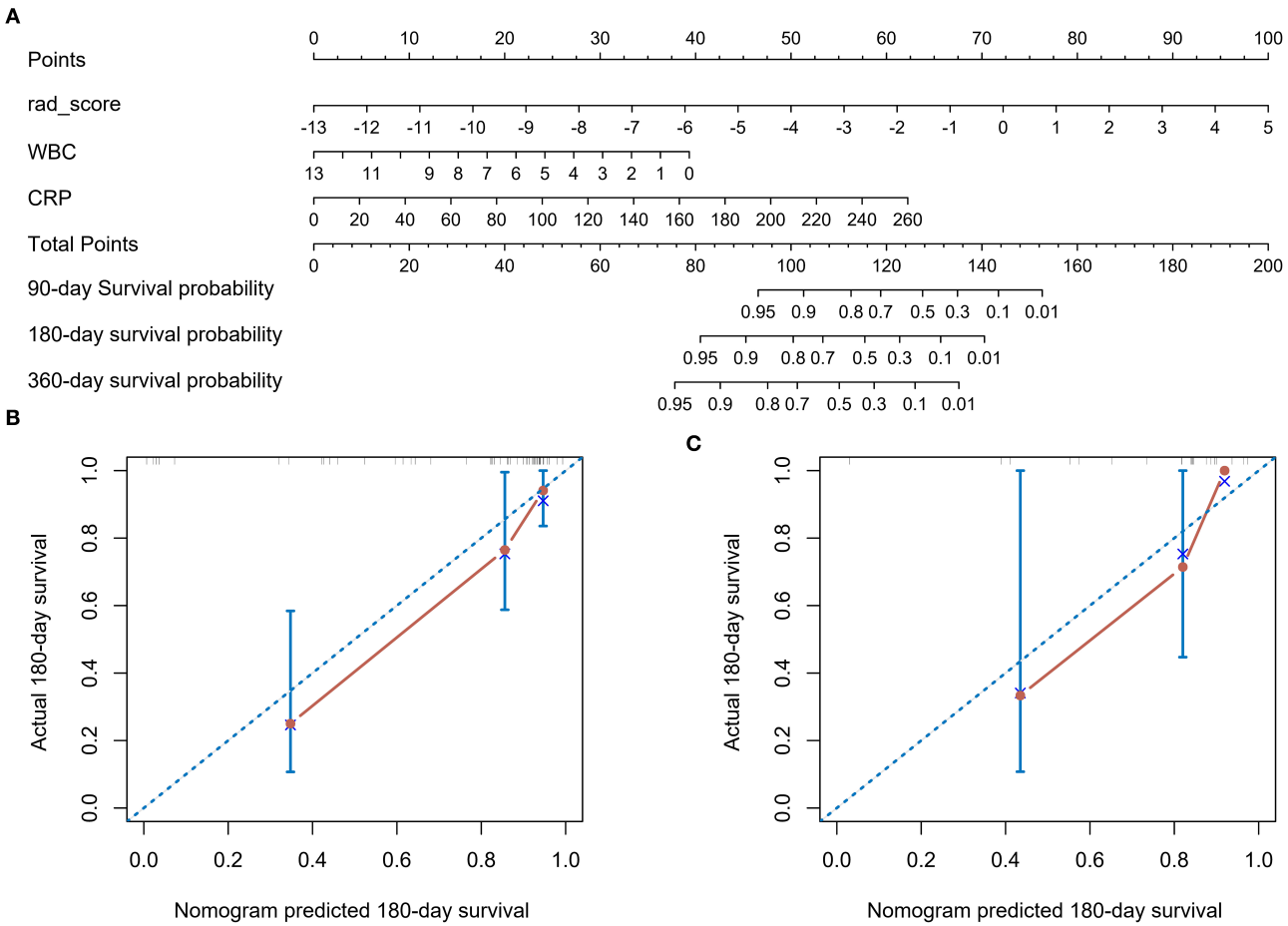


FIGURE 5 | The radiomics nomogram for predicting overall survival for adult HLH patients **(A)**. Points for Rad-score, WBC, and CRP can be obtained by calibrating with the point caliper, and then combined to obtain a total score that can be calibrated with the patient's probability of survival at different time. Calibration curves of the radiomics nomogram in the training cohort **(B)** and validation cohort **(C)**. Nomogram-estimated OS is plotted on the x-axis; the observed OS is plotted on the y-axis. The diagonal dotted line is a perfect estimation by an ideal model, in which the estimated outcome perfectly corresponds to the actual outcome. The solid red line represents performance of the nomogram: A smaller distance of the scatter points from the dotted line indicates better calibration.

ALT, are identified as indicators of various diseases including HLH. High ALT and AST/ALT ratio have been found to act as adverse prognostic factors in adult HLH (8, 28). Furthermore, hepatic involvement and hepatomegaly reveal poor prognostic indicators and early death predictors in HLH (8, 28). In line with these studies, our results illustrated that the radiomic features of spleen and liver presented great prognostic values for adult HLH.

Radiomics is a high-throughput extraction of quantitative information from medical images as well as subtle manifestations that are difficult to recognize or quantify by human eyes. Compared with the traditional PET/CT metrics, the radiomic features may reflect the pathological process much more sufficiently in the spleen and liver of patients with HLH. In this study, the majority of the selected radiomic features (5/6) were derived from wavelet decomposition images, indicating that wavelet transforms emphasize image details. It is very likely that wavelet decomposition images contain inconspicuous prognostic information (29, 30). GLSZM quantifies the number of groups of interconnected neighboring voxels with the same gray level intensity. NGTDM represents contrast, quantifying the difference between the gray level of a voxel and the average level of its neighbors within a distance. GLCM captures spatial relationships of pairs of voxels, while kurtosis is a first-order feature expressing the peak of the distribution of values in the ROI. All the selected features describe the texture of the spleen and liver quantitatively, reflecting uniformity or heterogeneity in both the organs (31). Previous studies found that intratumor heterogeneity was associated with poor outcome in various malignancies (15, 17, 30, 32, 33). HLH is a heterogeneous disease with various etiological components and complex underlying genetic variant types (3). Each possible etiology has distinct clinical characteristics and prognosis. Even in lymphoma-associated HLH that has the worst prognosis, the treatment

response is diverse (34). HLH could occur in EBV-associated T-/NK-cell lymphoproliferative disorders, which is a spectrum of disease from infection to malignancy. The histological features and immunophenotype are markedly heterogeneous. As in children, multiple gene mutations are linked to the development of HLH in adults, especially with the EBV-driven lymphoma, which requires hematopoietic cell transplant (6, 35, 36). The ^{18}F-FDG appearance of the liver and spleen came in various sizes, densities, and metabolisms, which reflected the heterogeneity of HLH. Radiomics quantified the spatial complexity of them. This study exhibited that the heterogeneity of spleen and liver may reveal overproliferation of immune cells accompanied with inflammatory infiltration triggered by EBV infection (37–40). On the other hand, the heterogeneous distribution of metabolism or density may also suggest the involvement of tumor cells (41–43). Both the malignancy and EBV infection seemed to link an inferior prognosis in adult HLH (9, 34, 44, 45). Additionally, the radiomic features have the possibilities associated with genetic signatures (3, 46); however, the underlying biological significance of these radiomic features has not been fully studied and the relationships among radiomic features, genetic signatures, and prognosis need further exploration in HLH.

Recent studies pointed out that a number of clinical parameters play a role in the prognosis of HLH such as lymphoid malignancy, hemoglobin, platelets, CRP, and cytopenia [(8, 9, 50, 51)]. It is well-documented that lymphoid malignancy is negatively associated with survival. Typically, T-cell lymphoma is acknowledged to have a more severe survival due to poor response to chemotherapy, in comparison with B-cell lymphoma (47). In a large-scale Japanese study, the 5-year OS was the worst in T-/NK-cell lymphoma-associated HLH compared with other types of HLH including primary HLH, B-cell lymphoma-associated HLH, and infection-associated HLH (48). Lower hemoglobin and platelet have been reported to be the more consistent negative prognostic biomarkers in HLH (8, 49). This study consistently showed that these 3 clinical parameters were involved in the clinical prediction model. However, the two clinical variables incorporated in the nomogram were white blood cell and CRP. Cytopenia is one of the major presentation in HLH. Serious cytopenia may mark the severity of a cytokine storm and lead to hemorrhage and sepsis, suggesting to be an inferior factor (7). CRP is the prototypical acute phase serum protein, increasing rapidly during inflammation (50). It has been highlighted that CRP is markedly enhanced in patients with secondary HLH compared to primary ones. High CRP levels have been correlated with increased risk of infection and overall mortality in HLH, suggested to be indices of disease severity (51). CRP probably serves as a predictor of ^{18}F-FDG PET/CT effectiveness due to the fact that the diagnostic accuracy of PET/CT is positively linked with CRP $>$ 60 mg/l in HLH (12).

Interestingly, T-cell neoplasms were not retained in the predictive model when the Rad-score was incorporated. A possible explanation was that the Rad-score contained partial pathological information. The inclusion of the Rad-score not only improved the prognostic performance, but also simplified the prediction model. DCA demonstrated that the nomogram with the Rad-score and two clinical parameters was superior to the clinical model in terms of clinical application. Overall, the nomogram was successfully built to predict 3-month, 6-month, and 1-year survival of adults with HLH and the accuracy and clinical applicability of the model were verified through C-index, calibration curve, and DCA.

This study has several limitations. First, patients may have been missed for inclusion in a single-center study and selection bias may occur because of the retrospective nature of the study design. Second, the heterogeneity of the patients and treatments may affect our results. Third, gene, transcript, and protein signatures become increasingly important for the prognosis of adult HLH (3), but these data were not collected. Finally, the Rad-score was calculated using ROIs that were manually delineated in 3D slicer. It was time-consuming and inconvenient for clinical practice, so automatic or semi-automatic image segmentation will be needed. Notably, a multicenter and prospective study with larger cohort will be required to validate our findings in the future.

CONCLUSION

This preliminary study indicated that the pretherapeutic ^{18}F-FDG PET/CT radiomic features of spleen and liver are independent prognostic factors in adult HLH, with the heterogeneity of spleen and liver associated with inferior prognosis. Integrating radiomic features with clinical parameters show synergetic power for 6-month survival prediction compared to other models with radiomics features or clinical parameters alone. The nomogram has great potential for predicting individualized 3-month, 6-month, and 1-year survival, which may timely guide personalized treatments for adult HLH.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of Beijing Friendship Hospital of Capital Medical University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. No potentially identifiable human images or data are presented in the manuscript.

AUTHOR CONTRIBUTIONS

JY, YK, and HZ contributed to the study design, decision-making, and coordination of the study. XY, JLiu, XL, WW, and SZ contributed to the management of registration of cases

and collected PET/CT image data. XY, JLiu, XL, WW, and YK contributed to the image quality control, analysis, and data interpretation. LL and HZ contributed to the statistical analysis. XY, JLi, and JY contributed to the drafting and revising the manuscript. All the authors read, revised, and approved the final version of the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Nos: 81971642 and 81771860), the Beijing Natural Science Foundation (No: 7192041), and the National Key Research and Development Plan (No: 2020YFC0122000).

REFERENCES

1. Al-Samkari H, Berliner N. Hemophagocytic lymphohistiocytosis. *Annu Rev Pathol.* (2018) 13:27–49. doi: 10.1146/annurev-pathol-020117-043625
2. Birndt S, Schenk T, Heinevetter B, Brunkhorst FM, Maschmeyer G, Rothmann F, et al. Hemophagocytic lymphohistiocytosis in adults: collaborative analysis of 137 cases of a nationwide German registry. *J Cancer Res Clin Oncol.* (2020) 146:1065–77. doi: 10.1007/s00432-020-03139-4
3. Zhang J, Sun Y, Shi X, Zhang R, Wang Y, Xiao J, et al. Genotype characteristics and immunological indicator evaluation of 311 hemophagocytic lymphohistiocytosis cases in China. *Orphanet J Rare Dis.* (2020) 15:112. doi: 10.1186/s13023-020-01390-z
4. Canna SW, Marsh RA. Pediatric hemophagocytic lymphohistiocytosis. *Blood.* (2020) 135:1332–43. doi: 10.1182/blood.2019000936
5. Ramos-Casals M, Brito-Zeron P, Lopez-Guillermo A, Khamashta MA, Bosch X. Adult haemophagocytic syndrome. *Lancet.* (2014) 383:1503–16. doi: 10.1016/S0140-6736(13)61048-X
6. La Rosée P, Horne A, Hines M, von Bahr Greenwood T, Machowicz R, Berliner N, et al. Recommendations for the management of hemophagocytic lymphohistiocytosis in adults. *Blood.* (2019) 133:2465–77. doi: 10.1182/blood.2018894618
7. Merrill SA, Naik R, Streiff MB, Shanbhag S, Lanzkron S, Braunstein EM, et al. A prospective quality improvement initiative in adult hemophagocytic lymphohistiocytosis to improve testing and a framework to facilitate trigger identification and mitigate hemorrhage from retrospective analysis. *Medicine (Baltimore).* (2018) 97:e11579. doi: 10.1097/MD.0000000000011579
8. Zhou J, Zhou J, Wu ZQ, Goyal H, Xu HG. A novel prognostic model for adult patients with hemophagocytic lymphohistiocytosis. *Orphanet J Rare Dis.* (2020) 15:215. doi: 10.1186/s13023-020-01496-4
9. Yoon SE, Eun Y, Huh K, Chung CR, Yoo IY, Cho J, et al. A comprehensive analysis of adult patients with secondary hemophagocytic lymphohistiocytosis: a prospective cohort study. *Ann Hematol.* (2020) 99:2095–104. doi: 10.1007/s00277-020-04083-6
10. Yuan L, Kan Y, Meeks JK, Ma D, Yang J. 18F-FDG PET/CT for identifying the potential causes and extent of secondary hemophagocytic lymphohistiocytosis. *Diagn Interv Radiol.* (2016) 22:471–5. doi: 10.5152/dir.2016.15226
11. Kim J, Yoo SW, Kang SR, Bom HS, Song HC, Min JJ. Clinical implication of F-18 FDG PET/CT in patients with secondary hemophagocytic lymphohistiocytosis. *Ann Hematol.* (2014) 93:661–7. doi: 10.1007/s00277-013-1906-y
12. Zheng Y, Hu G, Liu Y, Ma Y, Dang Y, Li F, et al. The role of (18)F-FDG PET/CT in the management of patients with secondary haemophagocytic lymphohistiocytosis. *Clin Radiol.* (2016) 71:1248–54. doi: 10.1016/j.crad.2016.05.011
13. Shieh AC, Guler E, Smith DA, Tirumani SH, Beck RC, Ramaiya NH. Hemophagocytic lymphohistiocytosis: a primer for radiologists. *AJR Am J Roentgenol.* (2020) 214:W11–W9. doi: 10.2214/AJR.19.21788
14. Abou Shaar R, Eby CS, van Dorp S, de Witte T, Otrock ZK. Increasing ferritin predicts early death in adult hemophagocytic lymphohistiocytosis. *Int J Lab Hematol.* (2021) 43:1024–31. doi: 10.1111/ijlh.13489
15. Wang H, Zhao S, Li L, Tian R. Development and validation of an (18)F-FDG PET radiomic model for prognosis prediction in patients with nasal-type extranodal natural killer/T cell lymphoma. *Eur Radiol.* (2020) 30:5578–87. doi: 10.1007/s00330-020-06943-1
16. Cottereau AS, Nioche C, Dirand AS, Clerc J, Morschhauser F, Casasnovas O, et al. (18)F-FDG PET dissemination features in diffuse large B-cell lymphoma are predictive of outcome. *J Nucl Med.* (2020) 61:40–5. doi: 10.2967/jnumed.119.229450
17. Arshad MA, Thornton A, Lu H, Tam H, Wallitt K, Rodgers N, et al. Discovery of pre-therapy 2-deoxy-2-(18)F-fluoro-D-glucose positron emission tomography-based radiomics classifiers of survival outcome in non-small-cell lung cancer patients. *Eur J Nucl Med Mol Imaging.* (2019) 46:455–66. doi: 10.1007/s00259-018-4139-4
18. Henter JI, Horne A, Arico M, Egeler RM, Filipovich AH, Imashuku S, et al. HLH-2004: diagnostic and therapeutic guidelines for hemophagocytic lymphohistiocytosis. *Pediatr Blood Cancer.* (2007) 48:124–31. doi: 10.1002/pbc.21039
19. Kikinis R, Pieper SD, Vosburgh KG. 3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support. In: Jolesz FA, editor. *Intraoperative Imaging and Image-Guided Therapy*. New York, NY: Springer New York (2014). p. 277-89. doi: 10.1007/978-1-4614-7657-3_19
20. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* (2017) 77:e104–e7. doi: 10.1158/0008-5472.CAN-17-0339
21. Liu J, Yang X, Yang J. Prognosis predicting value of semiquantitative parameters of visceral adipose tissue and subcutaneous adipose tissue of (18)F-FDG PET/CT in newly diagnosed secondary hemophagocytic lymphohistiocytosis. *Ann Nucl Med.* (2021) 35:386–96. doi: 10.1007/s12149-021-01577-9
22. Yang YQ, Ding CY, Xu J, Fan L, Wang L, Tian T, et al. Exploring the role of bone marrow increased FDG uptake on PET/CT in patients with lymphoma-associated hemophagocytic lymphohistiocytosis: a reflection of bone marrow involvement or cytokine storm? *Leuk Lymphoma.* (2016) 57:291–8. doi: 10.3109/10428194.2015.1048442
23. Bronte V, Pittet MJ. The spleen in local and systemic regulation of immunity. *Immunity.* (2013) 39:806–18. doi: 10.1016/j.immuni.2013.10.010
24. Pijl JP, Kwee TC, Slart R, Yakar D, Wouthuyzen-Bakker M, Glaudemans A. Clinical implications of increased uptake in bone marrow and spleen on FDG-PET in patients with bacteremia. *Eur J Nucl Med Mol Imaging.* (2020) 48:1467–77. doi: 10.1007/s00259-020-05071-8
25. Kalkanis A, Kalkanis D, Drougas D, Vavougios GD, Datseris I, Judson MA, et al. Correlation of spleen metabolism assessed by 18F-FDG PET with serum interleukin-2 receptor levels and other biomarkers in patients with untreated sarcoidosis. *Nucl Med Commun.* (2016) 37:273–7. doi: 10.1097/MNM.0000000000000431
26. Ahn SS, Hwang SH, Jung SM, Lee SW, Park YB, Yun M, et al. Evaluation of spleen glucose metabolism using (18)F-FDG PET/CT in patients with febrile autoimmune disease. *J Nucl Med.* (2017) 58:507–13. doi: 10.2967/jnumed.116.180729
27. Wei J, Jiang H, Gu D, Niu M, Fu F, Han Y, et al. Radiomics in liver diseases: current progress and future opportunities. *Liver Int.* (2020) 40:2050–63. doi: 10.1111/liv.14555
28. Yin G, Man C, Liao S, Qiu H. The prognosis role of AST/ALT (De Ritis) ratio in patients with adult secondary hemophagocytic lymphohistiocytosis. *Mediators Inflamm.* (2020) 2020:5719751. doi: 10.1155/2020/5719751
29. Chaddad A, Daniel P, Niazi T. Radiomics evaluation of histological heterogeneity using multiscale textures derived from 3D wavelet transformation of multispectral images. *Front Oncol.* (2018) 8:96. doi: 10.3389/fonc.2018.00096
30. Jiang Y, Yuan Q, Lv W, Xi S, Huang W, Sun Z, et al. Radiomic signature of (18)F fluorodeoxyglucose PET/CT for prediction of gastric cancer survival and chemotherapeutic benefits. *Theranostics.* (2018) 8:5915–28. doi: 10.7150/thno.28018

31. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to radiomics. *J Nucl Med.* (2020) 61:488–95. doi: 10.2967/jnumed.118.222893
32. Lv W, Yuan Q, Wang Q, Ma J, Feng Q, Chen W, et al. Radiomics analysis of PET and CT components of PET/CT imaging integrated with clinical parameters: application to prognosis for nasopharyngeal carcinoma. *Mol Imaging Biol.* (2019) 21:954–64. doi: 10.1007/s11307-018-01304-3
33. Lue KH, Wu YF, Liu SH, Hsieh TC, Chuang KS, Lin HH, et al. Prognostic value of pretreatment radiomic features of 18F-FDG PET in patients with hodgkin lymphoma. *Clin Nucl Med.* (2019) 44:e559–e65. doi: 10.1097/RLU.0000000000002732
34. Li B, Guo J, Li T, Gu J, Zeng C, Xiao M, et al. Clinical characteristics of hemophagocytic lymphohistiocytosis associated with non-Hodgkin B-cell lymphoma: a multicenter retrospective study. *Clin Lymphoma Myeloma Leuk.* (2021) 21:e198–205. doi: 10.1016/S2152-2650(21)02106-6
35. Jin Z, Wang Y, Wei N, Wang Z. Adult primary hemophagocytic lymphohistocytosis associated with lymphoma. *Ann Hematol.* (2020) 99:663–5. doi: 10.1007/s00277-020-03924-8
36. Ghosh S, Köstel Bal S, Edwards ESJ, Pillay B, Jiménez Heredia R, Erol Cipe F, et al. Extended clinical and immunological phenotype and transplant outcome in CD27 and CD70 deficiency. *Blood.* (2020) 136:2638–55. doi: 10.1182/blood.2020006738
37. Lu J, Fang Q, Ma C, Su F, Chen G, Huang M, et al. Atypical Epstein-Barr virus-associated hemophagocytic lymphohistiocytosis simulating lymphadenitis on (18)F-FDG PET/CT and its differential diagnosis. *Hell J Nucl Med.* (2017) 20:254–7. doi: 10.1967/s002449910612
38. Hao R, Yang X, Liu Z, Yang J. EBV-Associated T-cell lymphoproliferative disorders demonstrated on FDG PET/CT in a patient with hemophagocytic lymphohistiocytosis. *Clin Nucl Med.* (2019) 44:829–30. doi: 10.1097/RLU.0000000000002649
39. Pan Q, Luo Y, Wu H, Ma Y, Li F. Epstein-Barr Virus-associated hemophagocytic lymphohistiocytosis mimicking lymphoma on FDG PET/CT. *Clin Nucl Med.* (2018) 43:125–7. doi: 10.1097/RLU.0000000000001923
40. Thomas DL, Syrbu S, Graham MM. Epstein-Barr virus mimicking lymphoma on FDG-PET/CT. *Clin Nucl Med.* (2009) 34:891–3. doi: 10.1097/RLU.0b013e3181bed135
41. Suga K, Kawakami Y, Hiyama A, Matsunaga N, Imoto S, Fukuda N, et al. F-18 FDG PET/CT findings in a case of T-cell lymphoma-associated hemophagocytic syndrome with liver involvement. *Clin Nucl Med.* (2010) 35:116–20. doi: 10.1097/RLU.0b013e3181c7bf20
42. Harada S, Shinohara T, Naruse K, Machida H. Diffuse 18F-fluorodeoxyglucose accumulation in the bone marrow of a patient with haemophagocytic lymphohistiocytosis due to Hodgkin lymphoma. *BMJ Case Rep.* (2016) 2016:bcr2016217555. doi: 10.1136/bcr-2016-217555
43. Donald JS, Barnthouse N, Chen DL. Rare variant of intravascular large B-cell lymphoma with hemophagocytic syndrome. *Clin Nucl Med.* (2018) 43:e125–e6. doi: 10.1097/RLU.0000000000001969
44. Otrock ZK, Eby CS. Clinical characteristics, prognostic factors, and outcomes of adult patients with hemophagocytic lymphohistiocytosis. *Am J Hematol.* (2015) 90:220–4. doi: 10.1002/ajh.23911
45. Wang J, Wang D, Zhang Q, Duan L, Tian T, Zhang X, et al. The significance of pre-therapeutic F-18-FDG PET-CT in lymphoma-associated hemophagocytic lymphohistiocytosis when pathological evidence is unavailable. *J Cancer Res Clin Oncol.* (2016) 142:859–71. doi: 10.1007/s00432-015-2094-z
46. Tang Y, Tan H, Hu S. Is there any potential of FDG PET/CT in monitoring disease activity in familial hemophagocytic lymphohistiocytosis? *Clin Nucl Med.* (2018) 43:296–8. doi: 10.1097/RLU.0000000000002008
47. Jaffe ES. The 2008 WHO classification of lymphomas: implications for clinical practice and translational research. *Hematology Am Soc Hematol Educ Program.* (2009) 2009:523–31. doi: 10.1182/asheducation-2009.1.523
48. Ishii E, Ohga S, Imashuku S, Yasukawa M, Tsuda H, Miura I, et al. Nationwide survey of hemophagocytic lymphohistiocytosis in Japan. *Int J Hematol.* (2007) 86:58–65. doi: 10.1532/IJH97.07012
49. Pan H, Huo Y, Sun L. Comparison between clinical features and prognosis of malignancy- and non-malignancy-associated pediatric hemophagocytic lymphohistiocytosis. *BMC Pediatr.* (2019) 19:468. doi: 10.1186/s12887-019-1702-5
50. Marnell L, Mold C, Du Clos TW. C-reactive protein: ligands, receptors and role in inflammation. *Clin Immunol.* (2005) 117:104–11. doi: 10.1016/j.clim.2005.08.004
51. Ozen S, Dai A, Coskun E, Oztuzcu S, Ergun S, Aktekin E, et al. Importance of hyperbilirubinemia in differentiation of primary and secondary hemophagocytic lymphohistiocytosis in pediatric cases. *Mediterr J Hematol Infect Dis.* (2014) 6:e2014067. doi: 10.4084/mjhid.2014.067

Conflict of Interest: LL was employed by Sinounion Medical Technology (Beijing) Corporation, Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Medicine

ORIGINAL RESEARCH
published: 04 January 2022
doi: 10.3389/fmed.2021.792581

Single-Photon Emission Computed Tomography/Computed Tomography Image-Based Radiomics for Discriminating Vertebral Bone Metastases From Benign Bone Lesions in Patients With Tumors

Zhicheng Jin, Fang Zhang, Yizhen Wang, Aijuan Tian, Jianan Zhang, Meiyan Chen and Jing Yu**

Department of Nuclear Medicine, The Second Hospital of Dalian Medical University, Dalian, China

OPEN ACCESS

Edited by:
Chuantao Zuo,
Fudan University, China

Reviewed by:
Salvatore Annunziata,
Catholic University of the Sacred Heart, Italy
Virginia Liberini,
University of Turin, Italy

****Correspondence:***
Jing Yu
yujing_2020@dmu.edu.cn
Meiyan Chen
117075894@qq.com

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: *10 October 2021*
Accepted: *22 November 2021*
Published: *04 January 2022*

Citation:
Jin Z, Zhang F, Wang Y, Tian A, Zhang J, Chen M and Yu J (2022) Single-Photon Emission Computed Tomography/Computed Tomography Image-Based Radiomics for Discriminating Vertebral Bone Metastases From Benign Bone Lesions in Patients With Tumors. Front. Med. 8:792581. doi: 10.3389/fmed.2021.792581

Purpose: The purpose of this study was to investigate the feasibility of Single-Photon Emission Computed Tomography/Computed Tomography (SPECT/CT) image-based radiomics in differentiating bone metastases from benign bone lesions in patients with tumors.

Methods: A total of 192 lesions from 132 patients (134 in the training group, 58 in the validation group) diagnosed with vertebral bone metastases or benign bone lesions were enrolled. All images were evaluated and diagnosed independently by two physicians with more than 20 years of diagnostic experience for qualitative classification, the images were imported into MaZda software in Bitmap (BMP) format for feature extraction. All radiomics features were selected by least absolute shrinkage and selection operator (LASSO) regression and 10-fold cross-validation algorithms after the process of normalization and correlation analysis. Based on these selected features, two models were established: The CT model and SPECT model (radiomics features were derived from CT and SPECT images, respectively). In addition, a combination model (ComModel) combined CT and SPECT features was developed in order to better evaluate the predictive performance of radiomics models. Subsequently, the diagnostic performance between each model was separately evaluated by a confusion matrix.

Results: There were 12, 13, and 18 features contained within the CT, SPECT, and ComModel, respectively. The constructed radiomics models based on SPECT/CT images to discriminate between bone metastases and benign bone lesions not only had high diagnostic efficacy in the training group (AUC of 0.894, 0.914, 0.951 for CT model, SPECT model, and ComModel, respectively), but also performed well in the validation group (AUC; 0.844, 0.871, 0.926). The AUC value of the human experts was 0.849 and 0.839 in the training and validation groups, respectively. Furthermore, both SPECT model and ComModel show higher classification performance than human experts in the training group ($P = 0.021$ and $P = 0.001$, respectively) and the validation group

($P = 0.037$ and $P = 0.007$, respectively). All models showed better diagnostic accuracy than human experts in the training group and the validation group.

Conclusion: Radiomics derived from SPECT/CT images could effectively discriminate between bone metastases and benign bone lesions. This technique may be a new non-invasive way to help prevent unnecessary delays in diagnosis and a potential contribution in disease staging and treatment planning.

Keywords: radiomics, bone metastases, benign bone lesions, SPECT/CT, diagnosis

1. INTRODUCTION

Bone metastases were a common event in cancer evolution. Studies had shown that nearly 70% of cancer patients had metastases at autopsy, and 80% of the primary tumors were a prostate, breast, and lung cancers, bone-related events associated with bone metastases which can seriously affect patients' quality of life (1). Among patients with primary tumors with bone metastases or benign bone diseases, the early diagnosis was important for individualized patient treatment as treatment options vary widely (2). Although bone biopsy was the gold standard for identifying benign and malignant lesions, it was not widely used in clinical diagnosis and treatment because of the invasive procedure. A noninvasive method to distinguish bone metastases from benign bone lesions was urgently needed.

^{99m}Tc-labeled methylene diphosphonate (^{99m}Tc-MDP) whole-body scan (WBS) was frequently used in patients with bone lesions and had high sensitivity but low specificity. The radioactive tracer ^{99m}Tc-MDP was deposited in the bone by chemisorption and ion exchange, the abnormal uptake of the tracer reflected the osteogenic activity and local blood flow of the lesion (3). Single-Photon Emission Computed Tomography/Computed Tomography (SPECT/CT) combined anatomic and metabolic functions to improve the accuracy of anatomic localization of lesions and the specificity of bone imaging (4, 5). However, several researchers had indicated that bone metastases and benign bone lesions had similar imaging features, particularly for patients with already known cancer (6–8), it remained difficult to discriminate bone metastases and benign bone lesions as studies had shown that 14.3% of patients still had an equivocal diagnosis after SPECT/CT examination (9, 10). Moreover, SPECT/CT diagnosis mainly depended on physicians' personal experience, which inevitably had subjective factors, and it was difficult to quantify the intensity, uniformity, and heterogeneity of lesion distribution (11).

Radiomics convert digital images into mineable data through automated or semi-automatic and high-throughput methods. Radiomics could analyze the heterogeneity of tumors as a whole through hundreds of quantitative features and also analyze the quantitative relationship between tumor biological features and imaging features, which could construct models for tumor diagnosis, efficacy evaluation and prediction, and provide valuable references for clinical treatment of tumors (12, 13). The current research on radiomics mainly focused on CT and MRI (14–16). The pathological mechanism of bone metastasis was based on the disruption of the metabolic balance between osteoclasts and osteoblasts by the molecular action of cancer cells. In contrast, the benign bone disease showed inflammation and tissue remodeling of the periosteal cartilage tissue. Different osteoblastic and osteolytic mechanisms had the potential to cause different heterogeneity and distribution of radioactive tracer (17, 18). Furthermore, different from the anatomical information of the lesion provided by traditional imaging, SPECT/CT radiomics combined the anatomical information and metabolic information of the lesion to quantify the tumor heterogeneity, which had the potential to improve the diagnostic performance.

To the best of our knowledge, there were few studies related to bone diseases based on radiomics of SPECT/CT images. Therefore, the purpose of this study was to investigate the feasibility of SPECT/CT image-based radiomics in differentiating and improving diagnostic performance for bone metastases from benign bone lesions in patients with tumors.

2. MATERIALS AND METHODS

2.1. Patients

Participants between January 2019 and October 2020 were enrolled in this study according to the following inclusion criteria: 1) Patients a had history of the primary tumor; 2) Patients received SPECT/CT for further diagnosis because of abnormal uptake of vertebral radioactive tracer in 99mTc-MDP WBS; 3) Complete pathological, imaging, or clinical follow-up records and diagnosed with bone metastases or benign bone lesions; 4) At least one lesion in the spine and larger than 1 cm. In addition, the exclusion criteria included the following: 1) The shape of the lesion was irregular and difficult to delineate; 2) Abnormal uptake of radioactive tracer in SPECT images without lesions in CT images; 3) Had undergone surgery or medical treatment. The enrolled patients were randomly divided into the training and validation groups at the ratio of 7:3. The details of the participant's selection process were shown in **Figure 1**. This retrospective study was approved by the hospital ethics review committee and the requirement for informed patient consent was waived.

2.2. Image Acquisition

All the acquisition procedures were completed on SPECT/CT scanner equipped with a high resolution low energy parallel hole collimator (GE Healthcare Discovery NM/CT670 pro, USA). WBS was acquired within 2–5 h after intravenous administration of 15–25 mCi ^{99m}Tc-MDP (Beijing Atomic Hi-tech Co., LTD, China), then SPECT/CT was performed immediately for further

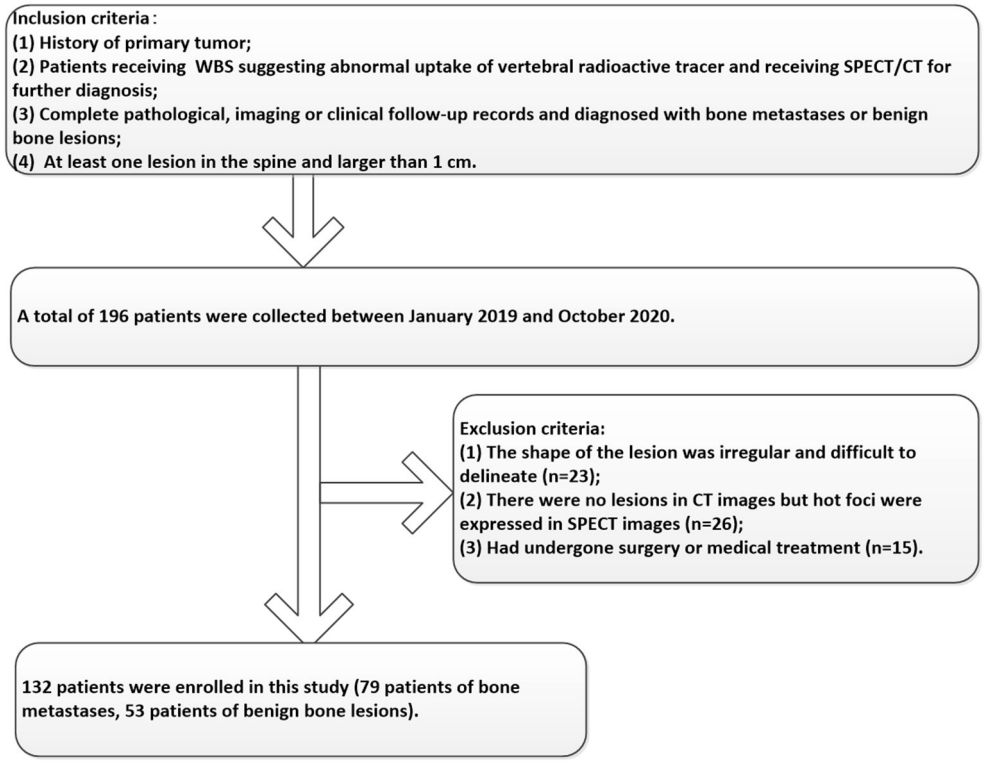


FIGURE 1 | The inclusion and exclusion criteria of our study.

diagnosis if a suspicious lesion was found on the WBS. The SPECT acquisition parameters were as follows: double probe parallel acquisition, rotation 180 respectively, 15 s per frame, and a 128×128 matrix. CT scan parameters are as follows: 120 KV, 80 mA, window width of 15%, the pitch of 1.25, and slice thickness of 2.5 mm. The image reconstruction program was carried out in the XELERIS workstation (GE Medical Systems, USA), and the image fusion program was carried out in the procedure of Volumetrix MI Evolution Bone.

2.3. Image Analysis and Human Expert's Qualitative Classification

After summarizing all the clinical information available for diagnosis, we concluded that the diagnostic criteria for this study were based on either pathological biopsy, follow-up imaging, or progression of the clinical course. All images were evaluated and diagnosed independently by two human experts (AJT and JY) with more than 20 years of diagnostic experience for qualitative classification. The human experts made the diagnosis without being provided with clinical information but were informed that the lesion was either bone metastasis or benign bone lesion. The diagnostic results of the human experts were evaluated by weighted kappa statistics for interobserver agreement. The main criteria for the human expert's qualitative classification of bone metastases were osteolytic, osteoblastic, and mixed bone changes on SPECT/CT images and abnormal uptake of ^{99m}Tc-MDP in the corresponding area.

2.4. Lesion Segmentation and Feature Extraction

All images were imported into MaZda software (version 4.6, www.eletel.p.lodz.pl) in BMP format for feature extraction, and at most two lesions were taken from each patient if the number of eligible lesions on the vertebral body was greater than three. MaZda software had been reported in previous studies to be available for radiomics image feature extraction, and it was confirmed that the radiomic features extracted by MaZda software satisfied the criteria of the Image Biomarker Standardization Initiative (IBSI) (19, 20). Before extracting features, images were normalized by using the method of $\mu \pm 3\sigma$ (μ is the average value of the image gray value, σ is the SD of the image gray value) to reduce the influence of brightness and contrast on the gray value of the image. Two physicians (MYC and JNZ) with 5 years of diagnostic experience checked the area of abnormal uptake of ^{99m}Tc-MDP as Region of Interest (ROI) without knowing the clear diagnosis of the lesion. The ROI was delineated on the largest cross-section of the lesion in CT and SPECT images using 2D texture sketching mode and then copied to corresponding images as needed. If the location of the lesion changes due to respiratory movement, the ROI was fine-tuned to ensure that the ROI is roughly in the same position. To

ensure consistency in outlining ROI between the two physicians and to maintain stability and reproducibility of the features, 30 lesions were randomly selected for secondary outlining. Features can be divided into the following categories: gray-level histogram, gray-level absolute gradient (GrM), gray-level run-length matrix (GLRLM), gray-level co-occurrence matrix (GLCM), autoregressive model (ARM), and wavelet. Detailed information about radiomics features had been explained in previous research (21). Altogether 279 radiomics parameters were included in 6 common feature groups.

2.5. Feature Selection and Model Establishment

All features were normalized using the method of Z-score (value of feature subtract the mean value and divided by the SD) before selection. We calculate inter-texture correlation by the method of Pearson correlation and remove features with a correlation coefficient greater than 0.9 to achieve data stability and repeatability as well as to eliminate the effect of multicollinearity. In our study, if the correlation coefficient between two features is greater than 0.9, the average absolute correlation between this feature is correlation coefficient and the remaining features was compared, and the feature coefficient with the larger correlation was removed. The least absolute shrinkage and selection operator (LASSO) regression was performed on the training group for further data selection. Then the features were selected by 10-fold cross-validation based on the criteria of binomial deviance minimization. For the final selected non-zero features, we constructed a classification model by the method of multiple logistic regression. Based on these selected features, two models were established: The CT model (texture parameters were derived from CT images only) and SPECT model (texture parameters were derived from SPECT images only). In addition, a combination model (ComModel) combined with CT and SPECT features was developed in order to better evaluate the predictive value of radiomics models. The flowchart of our study was shown in **Figure 2**.

2.6. Diagnostic Efficacy of Models and Comparison

The diagnostic efficacy of all models was evaluated by the area under the curve (AUC) of the receiver operating characteristic (ROC). The confusion matrix was used to calculate the overall accuracy of the models as well as the sensitivity, specificity, negative predictive value, and positive predictive value of each model. The DeLong test was used to compare the diagnostic efficacy between each model. Calibration curves and Brier score were used to evaluate the calibration of the categorical prediction models and the good of fitness. In addition, decision curve analysis (DCA) was used to evaluate the clinical benefit of the categorical prediction models.

2.7. Statistical Analysis

Descriptive data were represented as the mean ± SEM. Continuous variables were compared between groups of bone metastatic and benign bone lesions with the Independent-Samples *t*-Test or the Mann Whitney *U*-test for non-normal distribution. Categorical variables between the two groups were assessed using the chi-square test or the Fisher exact test and weighted Kappa statistics were used to evaluate the interobserver agreement. All feature screening, model construction, and evaluation of the radiomics model diagnostic efficacy were performed in R software (version 4.1.1) and Python (version 3.8.1). Other statistical analyses of clinical data were performed with IBM SPSS (version 21.0) and MedCalc software (version 20.0), and $P < 0.05$ was considered as statistically significant.

3. RESULTS

3.1. Basic Patient Information

A total of 192 lesions from 132 patients were enrolled in this study, which included 79 patients who were classified as bone metastasis (46 men, 33 women), while the remaining 53 patients were classified as benign bone lesions (32 men, 21 women). Among all of the basic clinical factors for patients in the training

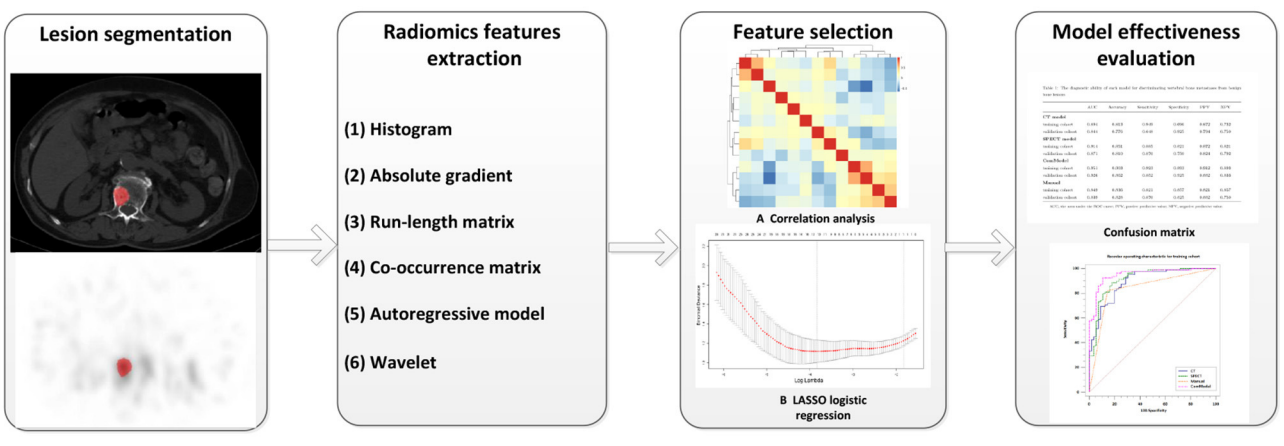


FIGURE 2 | The flowchart of our study.

TABLE 1 | Basic information for patients in the training and validation cohorts.

	The training cohort			The validation cohort		
	Bone metastases	**Benign lesions**	***P***	**Bone metastases**	**Benign lesions**	***P***
Gender			0.461			0.653
Female	19	10		13	11	
Male	35	26		11	7	
Age	59.13 ± 11.4	61.35 ± 12.32	0.508	58.46 ± 10.97	60.45 ± 11.65	0.632
Range	30–72	35–79		35–67	34–78	
Lesion form			0.839			0.801
Osteolytic	8	4		4	1	
Osteoblastic	53	24		23	11	
Mixed	33	12		14	5	
Confirmed			0.658			0.896
Biopsy	18	9		9	4	
Follow-up	76	31		32	13	

$P < 0.05$ was considered to be statistically significant.

and validation cohorts, including gender, age, lesion form, and methods of lesions confirmation showed no significant difference between bone metastases and benign bone lesion (all $P > 0.05$). The basic information of the patient was detailed in **Table 1**. The primary malignancies of the 132 patients were as follows: breast cancer, $n = 27$; lung cancer, $n = 50$; prostate cancer, $n = 25$; colon cancer, $n = 5$; renal carcinoma, $n = 5$; thyroid cancer, $n = 5$; stomach cancer, $n = 4$; cervical cancer, $n = 3$; hepatocellular cancer, $n = 3$; pancreatic cancer, $n = 2$; nasopharyngeal cancer, $n = 2$; ureteral cancer, $n = 1$. The diagnosis of the 53 patients with benign lesions is as follows: degenerative lesions, $n = 23$; fractures, $n = 15$; osteoarthritis, $n = 8$; spinal tuberculosis, $n = 7$.

3.2. Prediction Models Building and Validation

After correlation analysis between feature groups and elimination of features with a correlation greater than 0.9, 203, and 234 features were obtained from CT and SPECT images in the training group, respectively, and then the lasso algorithm and 10-fold cross-validation were used to classify bone metastases and benign bone lesions, and finally, 12, 13, and 18 features were obtained based on CT images and SPECT images for construction of classification models, respectively, the selected features were shown in **Supplementary Tables A,B**. The specific process of LASSO screening features was illustrated as detailed in **Figure 3**. The details of the selected features obtained were demonstrated with boxplots and heatmaps in **Supplementary Figures S1–S4**. In the training group, the CT model, the SPECT model, and the ComModel obtained high AUC values of 0.894 (95%CI: 0.829–0.941), 0.914 (95%CI: 0.853–0.956), and 0.951 (95%CI: 0.899–0.981), respectively. ComModel have better predictive performance than CT and SPECT and there was no statistical difference between the three models after DeLong test ($P = 0.622$ between SPECT model and CT model, $P = 0.193$ between SPECT and ComModel model, $P = 0.072$ between ComModel and CT model). In the validation group, ComModel (0.926; 95% CI: 0.827–0.978) indicted better predictive performance than SPECT model (0.871; 95% CI:0.757–0.945) and significant increase than CT model (0.844; 95% CI: 0.725–0.026) ($P = 0.063$ and $P = 0.024$, respectively). In addition, SPECT model also demonstrated better predictive performance than the CT model ($P = 0.042$).

3.3. Diagnostic Performance Between the CT Model, SPECT Model, ComModel, and Human Experts

After the Kappa test, the weighted *k*-value of the inter-observer agreement was 0.814 (95% CI: 0.713–0.895), indicating a good inter-observer agreement. The AUC value of the human experts' qualitative classification was 0.849 (95% CI: 0.775–0.907) and 0.839 (95% CI: 0.753–0.906) in the training and validation groups, respectively. In the training group, the SPECT model and the ComModel showed statistically significant differences over the human experts ($P = 0.021$ and $P = 0.001$, respectively), while the CT model showed no significant differences over the human experts ($P = 0.091$). In the validation group, the ComModel and SPECT model demonstrated greater diagnostic effectiveness over the human experts ($P = 0.007$ and $P = 0.037$ respectively), while the CT model showed no significant difference ($P = 0.094$). As for the calibration curves, all three model's curves were closed to ideal curves, indicating that the models had superior fitness and predictive ability. The calibration curve was shown in **Figure 4**. ComModel has a better model fitness than the CT model and SPECT model with a lower value of Brier score (0.082, 0.126, and 0.110 for ComModel, CT model and SPECT model, respectively). In the decision curves, when the threshold was 0–1, the ComModel always had a better overall net clinical gain than the other models. The SPECT model also had a slightly higher clinical gain than the human experts, and there was no significant difference between the CT model and the human experts. The decision curve was shown in **Figure 5**. The difficult differential diagnosis of bone metastases and benign bone lesions

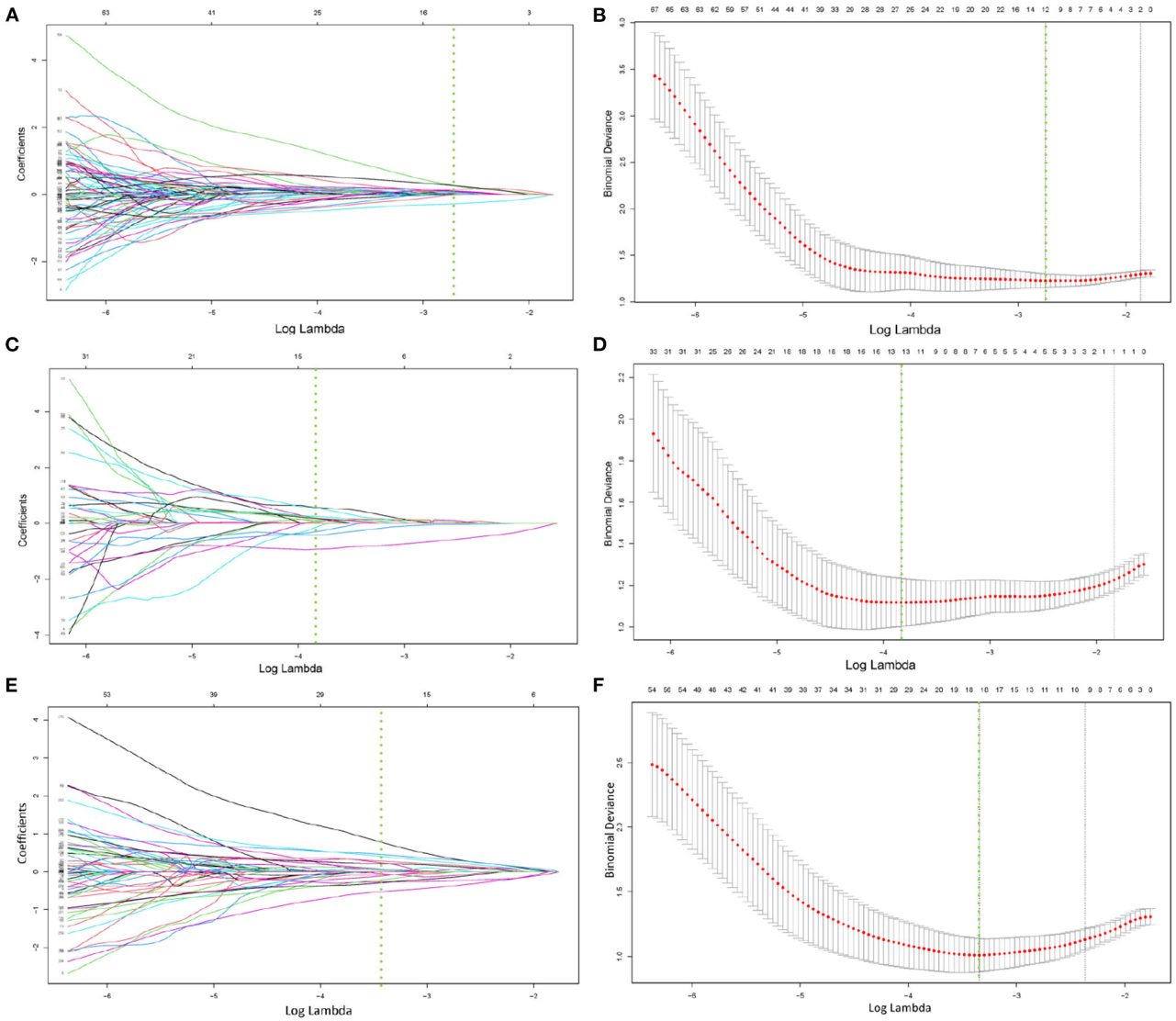


FIGURE 3 | (A–F) demonstrated the specific process of least absolute shrinkage and selection operator (LASSO) regression analysis screening features for CT model, SPECT model, and ComModel, respectively. **(A,C,E)** showed the process of features selection. The vertical line was plotted at the optimal λ of 0.064, 0.025, and 0.035 for CT, SPECT, and ComModel, respectively. Twelve, thirteen, and eighteen factors with non-zero coefficients were finally selected for CT, SPECT, and ComModel, respectively. **(B,D,F)** showed that features selection was performed by 10-fold cross-validation with the criterion of minimum deviance.

in clinical work was demonstrated in **Figures 6**, **7**, respectively. A comparison of diagnostic performance between each model was shown in **Table 2**. The ROC curves of all models were illustrated in **Figure 8**.

4. DISCUSSION

In this study, we constructed and validated SPECT/CT image-based radiomics model, which achieved satisfactory classification performance and outperformed human experts' qualitative classification. Radiomics model had the potential to provide additional value distinct from CT and MRI as a noninvasive and more accessible imaging method to differentiate bone metastases from benign bone disease and reduce unnecessary invasive examinations and adjustments in treatment decisions.

Previous studies had indicated that bone metastases tend to involve the pedicle rather than the vertebral body and rarely invade the extremity bone compared to benign bone disease, which tends to affect the small intervertebral joints (22). In a study of characterization of 84 solitary lesions in the extremities, Peng et al. (23) pointed out that benign bone lesions were predominant in the proximal and distal extremity bones, whereas bone metastases were predominant in the diaphyses extremity bones, but there was no significant difference in osteoblast activity between bone metastases and benign lesions (24).

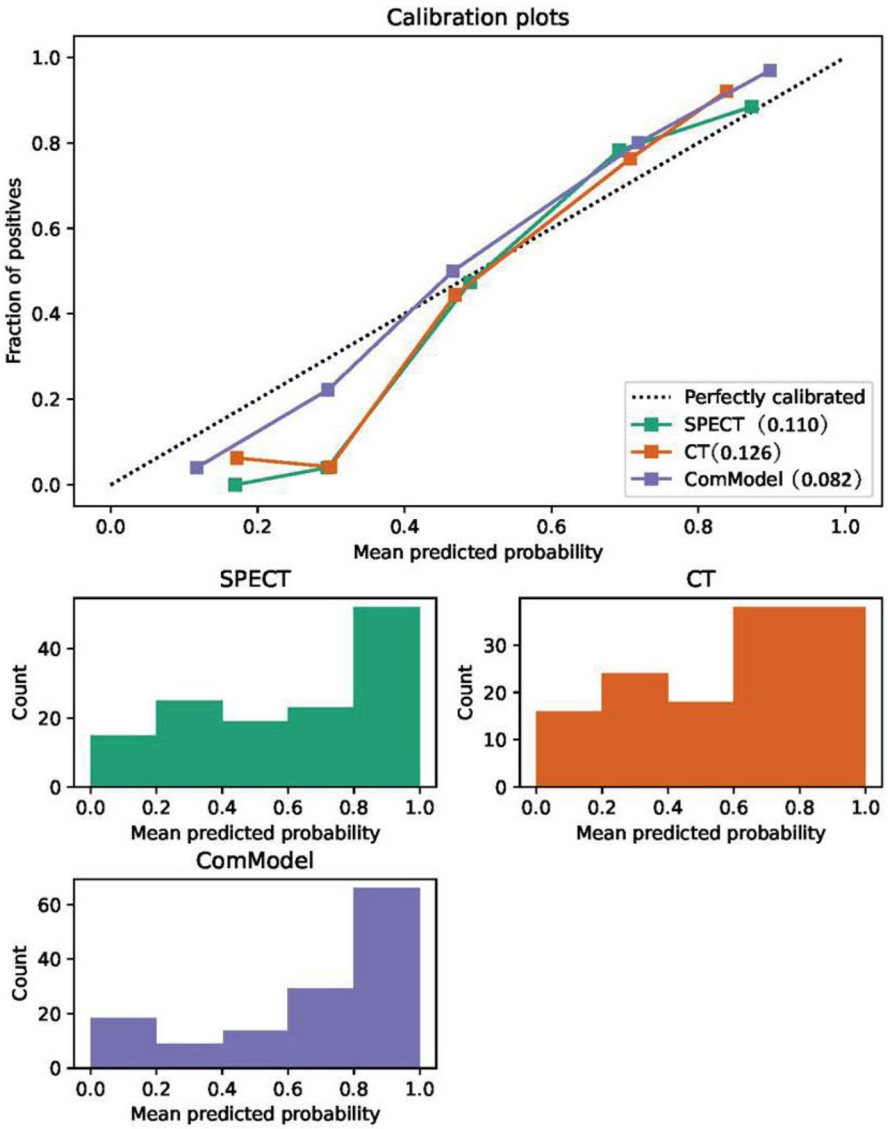


FIGURE 4 | Comparison of the calibration curve and Brier score of different models. All three model's calibration curves were closed to ideal curves, indicating that the models had good fitness and predictive ability. The following figure shows the distribution of the probability of diagnosis for different models.

Although SPECT/CT had significantly improved the diagnostic efficiency of spinal lesions and could diagnose bone metastases based on the criteria of osteolytic, osteoblastic, and mixed bone changes on SPECT/CT images and abnormal uptake of ^{99m}Tc-MDP in the corresponding area, some benign lesions such as fractures, degenerative changes, spinal tuberculosis, and osteoarthritis can also show similar bone changes in CT and abnormal uptake of radioactive tracer, furthermore, atypical bone lesions also contributed to the challenge of differentiating between the bone metastases and benign bone disease (25). In addition, these traditional imaging features were assessed through visualization and relied on the physician's subjective evaluation and diagnostic experience, despite that lesions were not always typical in the clinical work.

Considering the limitations of traditional imaging diagnosis, the semi-quantitative analysis of bone lesions had made great progress in recent years. Kuji et al. (26) used the method of conjugate gradient reconstruction with tissue zoning, attenuation, and scatter corrections applied (CGZAS) based on WBS image to prove that SUVmax is a reliable osteoblastic biomarker for differentiating bone metastasis from degenerative changes in patients with prostate cancer. In their study, SUVmax in patients with bone metastasis was significantly higher compared with degenerative changes (40.90 ± 33.46 vs. 16.73 ± 6.74). In addition, their study also showed that SUVmax was related to bone disease progression. Le et al. (27) showed that the differential of malignant bone metastases also achieved satisfactory diagnostic performance based on the factor of PSMA-RADS rating, SUVmax, and SUVmax ratio of the lesion to blood pool by 68Ga-PSMA-11 PET/CT image. In fact, SUVmax only reflects the metabolic information of the tumor within a single pixel in the image

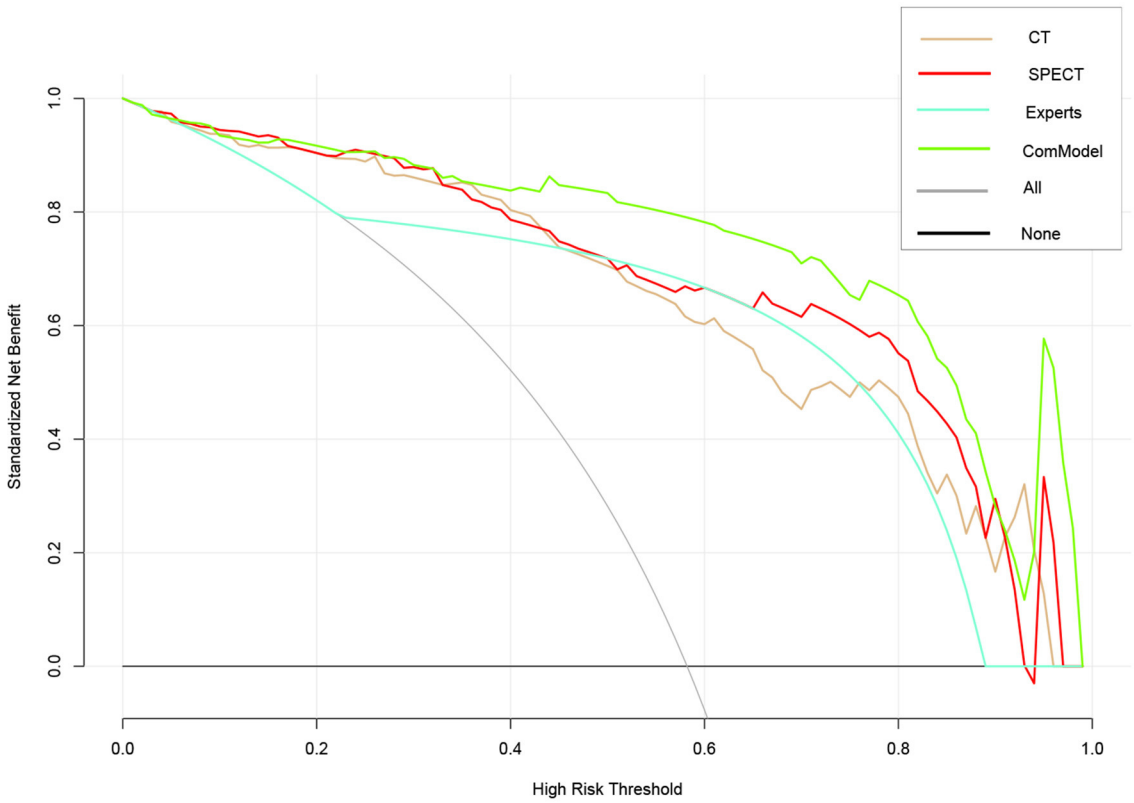


FIGURE 5 | Comparison of decision curve analysis (DCA) of different models. When the threshold was 0–1, the ComModel always had a better overall net clinical gain than the other models, the SPECT model also had higher clinical gain than the human experts, and there was no significant difference between the CT model and the human experts.

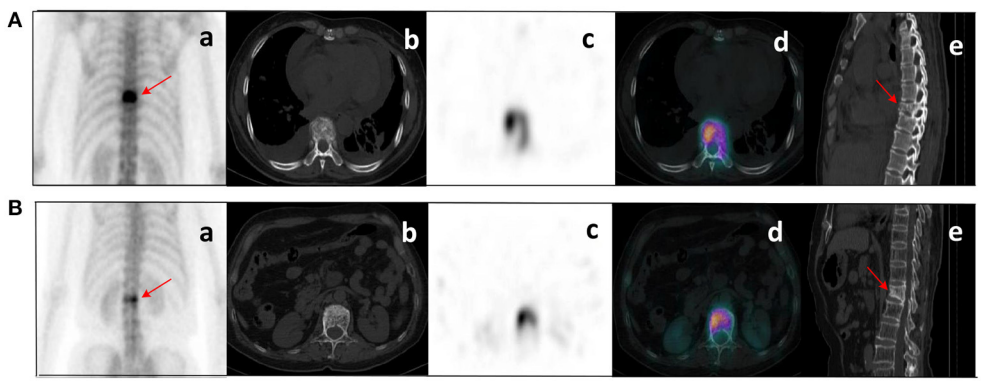


FIGURE 6 | Clinical cases SPECT/CT images of bone metastases **(A)** and benign bone lesions **(B)**. The images shown are WBS image, axial CT, SPECT, fusion image, and sagittal CT (a–e, respectively). **(A)** bone metastases: a 53-year-old female with an adenocarcinoma of the left lung. Wedge-like changes of the T8 vertebral body with an abnormal concentration of radioactive tracer (arrows). **(B)** benign bone lesions: a 68-year-old female with breast cancer. Wedge-like changes of the L1 vertebral body with higher bone density and increased radioactive tracer distribution (arrows). It was difficult to determine whether lesions were metastasis with conventional images only. Lesion **(A)** was confirmed as pathological fracture due to bone metastases by pathological examination and showed systemic bone metastases at subsequent imaging follow-up. Lesion **(B)** was confirmed to be a benign compression fracture by imaging follow-up and clinical information.

and cannot quantify the spatial heterogeneity of the overall metabolic distribution.

Compared to traditional image assessment, radiomics was a new tool that extracted image information through high-throughput methods to provide useful information for disease typing and grading, gene localization, early treatment, and prognostic assessment; Some studies had shown that radiomics had better diagnostic performance than traditional

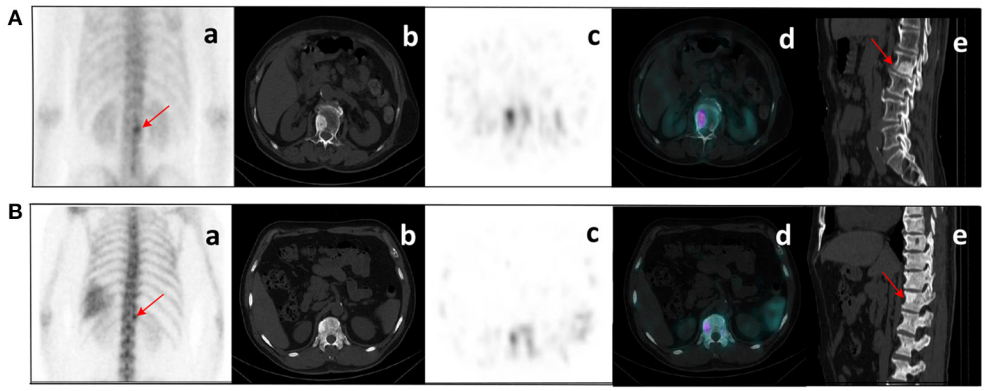


FIGURE 7 | (A) bone metastases: a 76-year-old male with prostate cancer. Nodular high-density shadow on the right lower edge of the L2 vertebral body with a concentrated radioactive tracer (arrows). **(B)** benign bone lesions: a 60-year-old male with prostate cancer. Nodular high-density shadow on the right upper edge of the T12 vertebral body with increased radioactive tracer distribution (arrows). The cluster of radioactive tracer concentrations in the left rib(a), with fusion images suggesting bone metastasis. Lesion **(A)** showed increased concentration of tracer and increased extent of concentration with systemic bone metastases at subsequent imaging follow-up. Lesion **(B)** was confirmed not a metastasis from prostate cancer at several subsequent imaging follow-ups.

TABLE 2 | The diagnostic ability of each model for discriminating vertebral bone metastases from benign bone lesions.

	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
CT model						
Training cohort	0.894	0.851	0.949	0.696	0.872	0.821
Validation cohort	0.844	0.828	0.648	0.925	0.853	0.792
SPECT model						
Training cohort	0.914	0.866	0.885	0.821	0.885	0.839
Validation cohort	0.871	0.845	0.870	0.750	0.853	0.833
ComModel						
Training cohort	0.951	0.903	0.923	0.893	0.912	0.893
Validation cohort	0.926	0.879	0.852	0.925	0.882	0.875
Human experts						
Training cohort	0.849	0.836	0.821	0.857	0.821	0.857
Validation cohort	0.839	0.828	0.870	0.825	0.882	0.750

AUC, the area under the ROC curve; PPV, postive predictive value; NPV, negative predictive value.

clinical in the non-invasive classification and diagnosis of diseases. Veres et al. (28) showed that SPECT radiomics could identify microscopic lesions in the rat liver and suggested that the radiomics feature skewness could identify liver tumor lesions before they exhibit altered tissue function. Carabelli et al. (29) demonstrated that the entropy of radiomic features in SPECT myocardial perfusion imaging (MPI) could evaluate coronary vascular microcirculation noninvasively and suggested that the improvement of left ventricular functional status by liraglutide would not improve the induction of coronary microvascular dysfunction in type 2 diabetes. The study of Rahmim et al. (30) combined with radiomic analysis in the routine measurement of DAT SPECT significantly improved the diagnosis of Parkinson's outcome and believed that radiomics were expected to become an effective biomarker for Parkinson's diagnosis. However, most of these SPECT radiomics studies focus on the brain and cardiovascular aspects, and there were few studies on bone diseases.

Our research showed that the radiomics models constructed based on SPECT/CT images to discriminate between bone metastases and benign bone lesions not only had high diagnostic efficacy in the training group, with AUC of 0.894, 0.914, and 0.951 for CT model, SPECT model, and ComModel, respectively but also performed well in the validation group, with AUC of 0.844, 0.871, and 0.926 for CT model, SPECT model, and ComModel, respectively. Furthermore, both SPECT model, and ComModel showed higher classification performance than human experts, which reflected the superiority of radiomics in non-invasive classification for disease diagnosis. Another finding was that the SPECT model had better diagnostic efficacy for identifying bone metastases and benign bone lesions than the CT model. We speculated that these SPECT images represent radioactive tracer uptake and metabolic information of the lesion and can detect the lesion earlier than conventional imaging, which brings additional value to the identification of the lesion and tissue specificity. In addition, SPECT radiomics may have the potential to play a

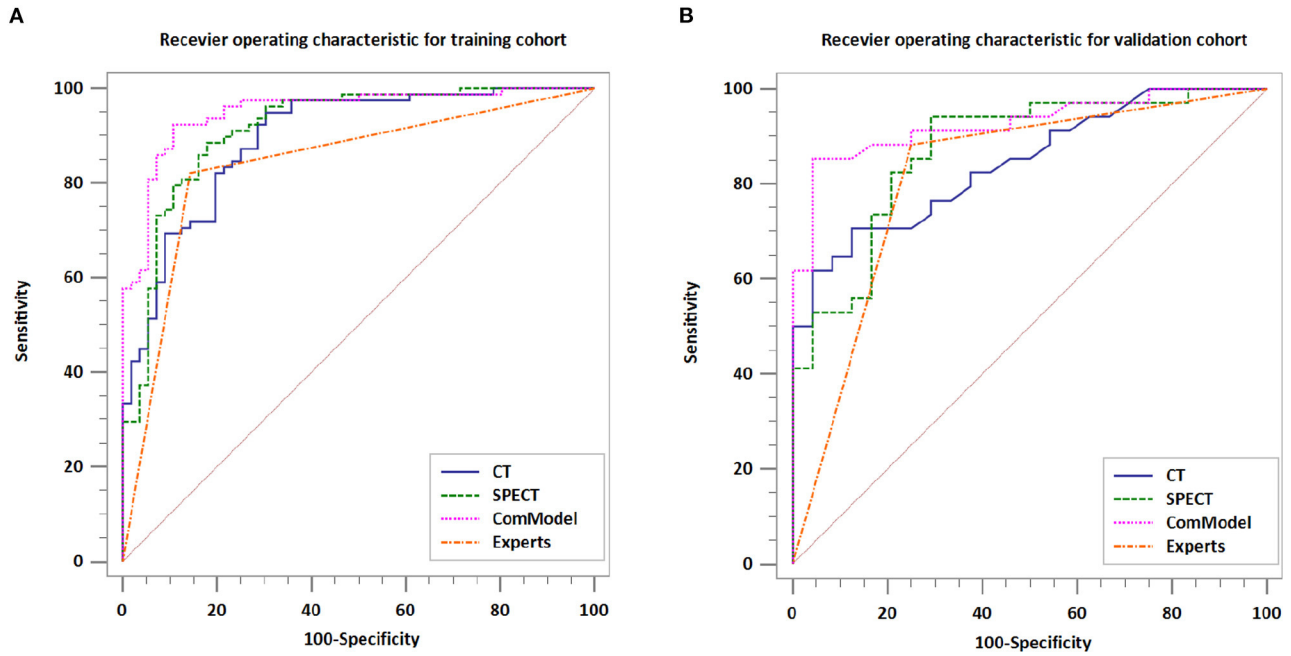


FIGURE 8 | Comparison of the diagnostic performance of different models. **(A)** Receiver operating characteristic for training cohort. **(B)** Receiver operating characteristic for validation cohort.

crucial role in finding the optimal dose for the targeted treatment of bone lesions with radioactive nuclides and in assessing the effectiveness of treatment.

Our study showed that the feature of entropy and correlation, which were both derived from GLCM and appeared several times in the feature extracted from CT images and SPECT images, were closely related to the identification of lesions. GLCM described the spatial relationship of pixels between features and the heterogeneity of lesions, which had been reported several times in previous studies (31, 32). Previous studies had found that the feature of entropy and correlation was related to the malignancy of lesions and helped to determine lymph node metastasis (33–35). In addition, increased heterogeneity within the images may be related to the local cellular composition, proliferation, fibrillation, angiogenesis, and necrosis of the tumor, as well as the impact of continued progressive invasion and destruction of bone metastatic disease (36, 37).

For patients considering multiple lesions, we do not automatically consider that all lesions in that patient were metastatic or benign based on biopsy or follow-up data of individual lesions, and the coexistence of bone metastases and benign bone disease in multiple lesions of the spine was relatively common in clinical work. We confirmed as metastases or benign lesions by biopsy or follow-up data for each individual lesion, and if the final diagnosis of the lesion was inconclusive, the lesion was simply eliminated, although this process took substantial time, it ensured the rigor of this study. Finally, we excluded treated patients because a proportion of patients will have flare phenomenon and osteoblastic reactions after chemotherapy or radiotherapy, which could also affect the uptake of radioactive tracer.

Our study has several limitations. First, our study had an inherent limitation with a retrospective design, thus, losing a large number of follow-up results, therefore, more standardized prospective studies are needed before the method can be used in the clinic. Second, our study is single-centered, and therefore, no external validation was performed, which may have some implications in terms of model stability. Third, detailed histopathological analysis was not always possible in each case, and we confirmed bone metastases and benign bone lesions on the basis of pathological biopsy, radiological imaging follow-up, and progression of the clinical course.

5. CONCLUSION

Radiomics models based on CT and SPECT images derived from SPECT/CT can effectively discriminate between vertebral bone metastases and benign bone disease. This technique may be a new non-invasive way to help prevent unnecessary delays in diagnosis and a potential contribution in disease staging and treatment planning.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

ZJ and JY contributed to the conception and design of the study. ZJ and MC organized the database. FZ and YW carried out data statistics and analysis. ZJ wrote the manuscript. JY and MC revised the manuscript. All authors have read and approved the final manuscript.

FUNDING

This study was supported by the Nature Science Foundation of Liaoning Province of China (No. 2019-ZD-0620).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.792581/full#supplementary-material

REFERENCES

1. Coleman RE. Metastatic bone disease: clinical features, pathophysiology and treatment strategies. *Cancer Treatment Rev*. (2001) 27:165–76. doi: 10.1053/ctrv.2000.0210
2. Coleman RE. Clinical features of metastatic bone disease and risk of skeletal morbidity. *Clin Cancer Res*. (2006) 12:6243. doi: 10.1158/1078-0432.CCR-06-0931
3. Kapsoritakis N, Stathaki M, Bourogianni O, Tsaroucha A, Papadaki E, Simos P, et al. Clinical impact of targeted single-photon emission computed tomography/computed tomography (SPECT/CT) bone scintigraphy on the assessment of bone metastasis in cancer patients. *Nuclear Med Commun*. (2021) 42:1202–8. doi: 10.1097/MNM.0000000000001455
4. Daisuke U, Shinya S, Masanori I, Seiji T, Koichi K, Shoji M, et al. Added value of SPECT/CT fusion in assessing suspected bone metastasis: comparison with scintigraphy alone and nonfused scintigraphy and CT. *Radiology*. (2016) 238:264–71. doi: 10.1148/radiol.2373041358
5. Helyar V, Mohan HK, Barwick T, Livieratos L, Gnanasegaran G, Clarke S, et al. The added value of multislice SPECT/CT in patients with equivocal bony metastasis from carcinoma of the prostate. *Eur J Nuclear Med Mol Imaging*. (2010) 37:706–13. doi: 10.1007/s00259-009-1334-3
6. Abikhzer G, Srour S, Keidar Z, Bar-Shalom R. Added value of SPECT/CT in the evaluation of benign bone diseases of the appendicular skeleton. *Clin Nuclear Med*. (2016) 41:e195. doi: 10.1097/RLU.0000000000001042
7. Zhang L, He Q, Li W, Zhang R. The value of 99mTc-methylene diphosphonate single photon emission computed tomography/computed tomography in diagnosis of fibrous dysplasia. *BMC Med Imaging*. (2017) 17:46. doi: 10.1186/s12880-017-0218-4
8. Tsai SY, Wang SY, Shiau YC, Wu YW. Benign incidental findings of osteopoikilosis on Tc-99m MDP bone SPECT/CT: a case report and literature review. *Medicine*. (2016) 95:e3868. doi: 10.1097/MD.0000000000003868
9. O'Sullivan G. Imaging of bone metastasis: an update. *World J Radiol*. (2015) 7:202. doi: 10.4329/wjr.v7.i8.202
10. Guezennec C, Keromnes N, Robin P, Abgral R. Incremental diagnostic utility of systematic double-bed SPECT/CT for bone scintigraphy in initial staging of cancer patients. *Cancer Imaging*. (2017) 17:16. doi: 10.1186/s40644-017-0118-4
11. Ritt P, Vija H, Hornegger J, Kuwert T. Absolute quantification in SPECT. *Eur J Nuclear Med Mol Imaging*. (2011) 38:69–77. doi: 10.1007/s00259-011-1770-8
12. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169
13. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RGPM, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
14. Marino MA, Pinker K, Leithner D, Sung J, Jochelson M. Contrast-enhanced mammography and radiomics analysis for noninvasive breast cancer characterization: initial results. *Mol Imaging Biol*. (2019) 22:780–7. doi: 10.1007/s11307-019-01423-5
15. Deng Y, Yang BR, Luo JW, Du GX, Luo LP. DTI-based radiomics signature for the detection of early diabetic kidney damage. *Abdom Radiol*. (2020) 45:2526–31. doi: 10.1007/s00261-020-02576-6
16. Cy A, Mh A, Sl B, Jc B, Yao YB, Na QB, et al. Radiomics model of magnetic resonance imaging for predicting pathological grading and lymph node metastases of extrahepatic cholangiocarcinoma. *Cancer Lett*. (2020) 470:1–7. doi: 10.1016/j.canlet.2019.11.036
17. Florimonte L, Dellavedova L, Maffioli LS. Radium-223 dichloride in clinical practice: a review. *Eur J Nuclear Med Mol Imaging*. (2016) 43:1896–909. doi: 10.1007/s00259-016-3386-5
18. Hayes AJ, Reynolds S, Nowell MA, Meakin LB, Habicher J, Ledin J, et al. Spinal deformity in aged zebrafish is accompanied by degenerative changes to their vertebrae that resemble osteoarthritis. *PLoS ONE*. (2013) 8:e75787. doi: 10.1371/journal.pone.0075787
19. Xh A, Jian SA, Yy A, Xin CA, Cy A, Tz B, et al. Feasibility of magnetic resonance imaging-based radiomics features for preoperative prediction of extrahepatic cholangiocarcinoma stage. *Eur J Cancer*. (2021) 155:227–35. doi: 10.1016/j.ejca.2021.06.053
20. Korte JC, Cardenas C, Hardcastle N, Kron T, Wang J, Bahig H, et al. Radiomics feature stability of open-source software evaluated on apparent diffusion coefficient maps in head and neck cancer. *Sci Rep*. (2021) 11:17633. doi: 10.1038/s41598-021-96600-4
21. Szczypiński P, Strzelecki M, Materka A, Klepaczko A. MaZda-A software package for image texture analysis. *Comput Methods Programs Biomed*. (2008) 94:66–76. doi: 10.1016/j.cmpb.2008.08.005
22. Even-Sapir E, Martin RH, Barnes DC, Pringle CR, Iles SE, Mitchell MJ. Role of SPECT in differentiating malignant from benign lesions in the lower thoracic and lumbar vertebrae. *Radiology*. (1993) 187:193. doi: 10.1148/radiology.187.1.8451412
23. Peng H, Zhang L, Zhou T, Li W, Li W, Ma L, et al. Characterization of solitary lesions in the extremities on whole-body bone scan in patients with known cancer: contribution of single-photon emission computed tomography/computed tomography. *Front Oncol*. (2007) 9:607. doi: 10.3389/fonc.2019.00607
24. Wedin R, Hansen BH, Laitinen M, Trovik C, Weiss RJ. Complications and survival after surgical treatment of 214 metastatic lesions of the humerus. *J Shoulder Elbow Surgery*. (2011) 21:1049–55. doi: 10.1016/j.jse.2011.06.019
25. Qian X, Wenqi Z, Shi G, Bin C, Qingjie M, Tianji L, et al. Indeterminate solitary vertebral lesions on planar scintigraphy. *Nuklearmedizin*. (2018) 57:216–23. doi: 10.3413/Nukmed-0973-18-04
26. Kuji I, Yamane T, Seto A, Yasumizu Y, Shirotake S, Oyama M. Skeletal standardized uptake values obtained by quantitative SPECT/CT as an osteoblastic biomarker for the discrimination of active bone metastasis in prostate cancer. *Eur J Hybrid Imaging*. (2017) 1:2. doi: 10.1186/s41824-017-0006-y
27. Le WC, Lawhn-Heath C, Behr SC, Juarez R, Flavell RR. Factors predicting metastatic disease in 68 Ga-PSMA-11 PET–positive osseous lesions in prostate cancer. *J Nuclear Med*. (2020) 61:jnumed.119.241174. doi: 10.2967/jnumed.119.241174

28. Veres DS, Mathe D, Hegedus N, Horvath I, Kiss FJ, Taba G, et al. Radiomic detection of microscopic tumorous lesions in small animal liver SPECT imaging. *EJNMMI Res*. (2019) 9:67. doi: 10.1186/s13550-019-0532-7
29. Carabelli A, Canu M, de Fondaumiere M, Debiossat M, Leenhardt J, Broisat A, et al. Noninvasive assessment of coronary microvascular dysfunction using SPECT myocardial perfusion imaging and myocardial perfusion entropy quantification in a rodent model of type 2 diabetes. *Eur J Nuclear Med Mol Imaging*. (2021) doi: 10.1007/s00259-021-05511-z. [Epub ahead of print].
30. Rahmim A, Huang P, Shenkov N, Fotouhi S, Davoodi BE, Lu L, et al. Improved prediction of outcome in Parkinson's disease using radiomics analysis of longitudinal DAT SPECT images. *NeuroImage Clin*. (2017) 16: 539–44. doi: 10.1016/j.nicl.2017.08.021
31. Huang Z, Mou L, He D, Wei Y, Yu H, Wang Y, et al. Two-dimensional texture analysis based on CT images to differentiate pancreatic lymphoma and pancreatic adenocarcinoma: a preliminary study. *Academic Radiol*. (2019) 26:S1076633218303969. doi: 10.1016/j.acra.2018.07.021
32. Feng M, Zhang M, Liu Y, Jiang N, Meng Q, Wang J, et al. Texture analysis of MR images to identify the differentiated degree in hepatocellular carcinoma: a retrospective study. *BMC Cancer*. (2020) 20:611. doi: 10.1186/s12885-020-07094-8
33. Zhu H, Xu Y, Liang N, Sun H, Wang W. Assessment of clinical stage IA lung adenocarcinoma with pN1/N2 metastasis using CT quantitative texture analysis. *Cancer Manag Res*. (2020) 2:6421–30. doi: 10.2147/CMAR.S251598
34. Beckers R, Trebeschi S, Maas M, Schnerr RS, Sijmons J, Beets GL, et al. CT texture analysis in colorectal liver metastases and the surrounding liver parenchyma and its potential as an imaging biomarker of disease aggressiveness, response and survival. *Eur J Radiol*. (2018) 102:15–21. doi: 10.1016/j.ejrad.2018.02.031
35. Moscoso A, Ruibal, Domínguez-Prado I, Fernández-Ferreiro A, Herranz M, Albaina L, et al. Texture analysis of high-resolution dedicated breast 18 F-FDG PET images correlates with immunohistochemical factors and subtype of breast cancer. *Eur J Nuclear Med Mol Imaging*. (2018) 45:196–206. doi: 10.1007/s00259-017-3830-1
36. Tixier F, Rest C, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nuclear Med*. (2011) 52:369–78. doi: 10.2967/jnumed.110.082404
37. Diessner J, Wischnewsky M, Stüber T, Stein R, Krockenberger M, Usler HS, et al. Evaluation of clinical parameters influencing the development of bone metastasis in breast cancer. *BMC Cancer*. (2016) 16:307. doi: 10.1186/s12885-016-2345-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership