Journal Articles

2020

Radiomics and supervised machine learning in the diagnosis of parkinsonism with FDG PET: promises and challenges.

S. Peng

P. G. Spetsieris

D. Eidelberg
Zucker School of Medicine at Hofstra/Northwell, deidelbe@northwell.edu

Y. Ma

Recommended Citation

Radiomics and supervised machine learning in the diagnosis of parkinsonism with FDG PET: promises and challenges

Shichun Peng, Phoebe G. Spetsieris, David Eidelberg, Yilong Ma

Center for Neurosciences, Institute of Molecular Medicine, The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, USA

Correspondence to: Yilong Ma, PhD. Center for Neurosciences, Institute of Molecular Imaging, The Feinstein Institutes for Medical Research, 350 Community Drive, Manhasset, New York 11030, USA. Email: yma@northwell.edu.

Radiomics is a converging multidisciplinary field of translational medicine aimed at automatic and completely data-centric diagnosis of any disease entity in individual patients using a variety of medical imaging technology and machine learning (ML) methodology. This usually entails the use of automated techniques for data processing and analysis followed by the application of powerful statistical solutions for predicting accurate diagnosis and prognosis on a single-subject basis. A wide array of computing methods have been developed for data reduction, feature extraction and classification. As a surrogate marker of synaptic activity, regional cerebral glucose metabolism measured with ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) has been one of most versatile and cost-effective neuroimaging biomarkers in the field of neurodegenerative disorders. Indeed, this biomarker has been widely used in the objective assessment of early differential diagnosis, clinical correlation, disease progression and therapeutic response in patients with Parkinson's disease (PD) as well as with atypical parkinsonism (APD) including multiple system atrophy (MSA), progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) (1-5). Both regional and brain network biomarkers are used in clinical and research applications of translational medicine based on mean and variance information of glucose metabolism provided by FDG PET. It is still a challenge to distinguish PD from any form of APD at early clinical stages despite the tremendous achievements over the last decade.

Discrimination of PD from healthy controls

A recent article on this journal reported a new method to discriminate early-stage patients with PD from normal control (NC) subjects using radiomic features extracted from brain FDG PET images (6). The work involved supervised feature selection based on texture analysis in a set of predetermined anatomical regions of interest alongside a classifier built with a support vector machine (SVM). The study taps into valuable diagnostic information from high-dimensional radiomic features that reflect regional brain tissue heterogeneity revealed by FDG PET. The diagnostic performance of this method was evaluated rigorously with bootstrap resampling where a large cohort of PD and NC subjects at one center were used as training and test sets with five-fold cross-validation, while a small cohort of PD and NC subjects at another center were used as a separate test set. Main outcomes were the accuracy, sensitivity and specificity averaged across multiple runs for each of the two test datasets. The authors showed high diagnostic accuracy in comparable patients with PD at early clinical stages versus age- and gender-matched healthy controls from two medical centers.

This study has put forth several major advances in the development of computer aided diagnosis for PD and related disorders with FDG PET. The authors selected 26 subcortical and cortical anatomical volumes of interest (VOIs) known to have abnormal metabolism from prior published studies. The matrix-based analysis resulted in

43 texture features, 188 wavelet features and 4 intensity features for each VOI. They then identified 30 independent radiomic features most relevant for disease discrimination by using autocorrelation and Fisher scoring algorithms. Of interest, the 10 top-ranked features were found across a set of relatively hypermetabolic regions of pons, cerebellum, pallidum and supplementary motor area (MSA) and relatively hypometabolic regions of inferior occipital cortex. In this study, three kernel (linear, sigmoid, and radial basis) functions were included to assess the ability of feature generalization and the reliability of SVM classification (6). The authors showed that diagnosis by the SVM classifier was similar among the three different kernels and improved slightly versus another classifier based on random forest. They further demonstrated with both classifiers that radiomic features provided superior classification compared to the traditional method that used features from mean metabolic values in these VOIs, and the combination of both types of features increased the diagnostic accuracy moderately. The latter is expected because some of these features were weakly correlated as predictive variables.

One important finding of this study is that some high-frequency radiomic features in the right pons, SMA and cerebellum correlated positively with clinical stages or severity of motor symptoms and also variably with corresponding metabolic values in these regions in the combined cohort of patients with PD. Thus, the top discriminatory radiomic features selected by the algorithms captured disease-related signatures despite their weak but significant associations with clinical variables and abnormal regional metabolism. This is also not surprising given the prior hypothesis concerning the VOIs included in the study as well as the strategy underling the extraction and selection of diagnostic features.

There are several areas of interest worth of further investigations on this topic. Firstly, decision scores produced for individual subjects by the SVM classifier appear to be a promising biomarker but this potential has to be fully evaluated with the same dataset for diagnostic use and clinical correlation. Secondly, it may be worthwhile to incorporate clinical stage and disease severity of patients in the ML models to improve the replicability across sites, in addition to the demographic information of individual subjects. Thirdly, it would be of importance to examine whether regionally-specific radiomic features have anything to do with clinical asymmetry of motor symptoms and subtype (i.e., akinesia, rigidity and resting tremor) rating scales in PD. Fourthly, the questions remain about whether

diagnostic performance can be further enhanced by using ratio images compared to those of standard uptake value (SUV) considering that the sensitivity for detecting group difference in regional glucose metabolism is often increased with FDG PET image ratio-normalized by mean value in a reference region like gray matter or white matter. Finally, although the proposed ML techniques marginally outperformed the diagnostic accuracy of deep belief learning network, it may be more valuable to compare with other innovative methods like deep convoluted neural network demonstrating superior accuracy for early diagnosis of PD versus NC subjects using a more specific biomarker of dopaminergic dysfunction (7).

One technical limitation of this study is that the large anatomical VOIs selected from the standard brain atlas are still primarily hypothesis-driven and may be less sensitive to localized metabolic alteration in PD. This was so despite the partial confirmation of these VOIs by the authors in comparison with brain mapping analysis in the large cohort of patients with PD and healthy controls (6). It would be necessary to determine whether diagnosis can be improved using radiomic features derived from more specific regions of abnormal glucose metabolism defined by voxel-wise brain mapping analysis with either univariate and multivariate models in neurodegenerative disorders (8,9). These can also offer disease-specific brain masks to help implement or refine any completely voxel-based ML algorithms for differential diagnosis.

Since early 2000s many studies with FDG PET images and voxel-based principal component analysis (PCA) have shown that individual patients with PD can be assessed reliably by elevated expression scores of a specific disease-related metabolic pattern called PDRP, which is highly reproducible across multiple imaging centers worldwide (5,10,11). It is worth noting that subject sores of PDRP demonstrated much stronger clinical correlations in early stage PD (11,12) than those with radiomic features as reported in the study by the authors (6). To this end it would be straight-forward to compare the performance of discriminating PD patients from healthy controls by using features obtained from subject scores of PDRP in the Chinese population (13) and determine whether this network biomarker could be incorporated in the predictive models to further increase the diagnostic accuracy. Another topic of growing interest was to examine whether the diagnostic methods proposed by the authors could help predict the onset of PD in prodromal patients with rapid eye movement sleep behavior disorder as reported

previously by subject scores of PDRP (14,15).

Early differential diagnosis

In clinical practice it is less a challenge to distinguish patients with PD from NC subjects by itself than from patients with APD. Typically, patients with uncertain diagnosis of parkinsonian disorders at early-stage consist of approximately 80% with PD and 20% with APD; some of whom may have never been treated with any antiparkinsonian medications. This study revealed relatively high performance in diagnosis given variable clinical stage and disease severity in patient cohorts across both centers (6). Nevertheless, the finding was limited by the absence of APD subjects and the retrospective nature of the study. The authors could test the specificity of their diagnostic methods by simply including APD subject as an additional test cohort.

Down the road the present studies could be extended to extract other disease-specific radiomic features associated with MSA, PSP and CBD. Diagnosis for each clinical phenotype of parkinsonism can be realized by a binary classification of PD versus APD and a multi-category classification among APD themselves. The primary premise of this approach is that diagnostic accuracy for PD would be higher by merely identifying and excluding APD. The performance of such a diagnosis at baseline must be confirmed prospectively until a final diagnosis is made after several years of clinical follow-up. Indeed, one study with FDG PET reported previously that the accuracy was high in discriminating PD from APD but poor for diagnosing each phenotype of APD with voxel-based features derived using relevance vector machine (16). By contrast, this strategy has produced superior differential diagnosis in terms of excellent specificity and positive predictive value even at early stage, by using multi-disease metabolic patterns as diagnostic features along with novel classification methods based on multivariate regression (1). The key input is subject scores computed in a large cohort of clinically undiagnosed parkinsonian subjects for metabolic brain networks associated specifically with PD, MSA and PSP. These disease-related brain networks had been identified previously in clinically established American patients by pattern recognition with supervised PCA based on brain-wide volumetric variance information of regional glucose metabolism (17,18). Subsequently, this robust method for differential diagnosis was cross-validated prospectively in a completely independent cohort in India (19), and proved

highly valuable in helping select PD patients for a successful multi-center clinical trial of gene therapy (20). It would be important to compare these methods with other related ML algorithms based on radiomic signatures extracted from independent component analysis (21) as well as partial least squares (22).

One major limitation commonly seen in the design of supervised ML algorithms is that training is mostly performed using homogeneous dataset at one site followed by testing using mixed dataset across multiple sites. More rigorous cross-validation in truly independent medical centers is warranted in light of differences in PET camera systems, imaging protocols, attenuation correction methods, image reconstruction algorithms, and any pre- or post-processing procedures (23,24). This will provide unique opportunity to refine ML models of learning and testing that can fully account for these technical factors as well as inevitable variability in patient populations and clinical expertise in multi-center settings. It would be necessary to evaluate whether prognostic outcomes can be improved by including baseline independent clinical variables such as disease profiles, medication status and chronic exposure.

Roadblock to clinical implementation

The work described in this article represents the important first step prior to implementing more comprehensive and rigorous radiomic approaches of this kind that can leverage novel methods for feature extraction and adaptive ML algorithms in the fields of artificial intelligence and biostatistics decision-making. It would be more challenging to extend the innovative methods put forward in this study in the context of early and differential diagnosis of parkinsonism. For instance, some of the top radiomic features selected by the authors may be highly variable and not replicable in prospective and independent validation across multiple centers (24). Any new techniques developed with this approach in the future should be compared with the established methods that have already shown great promise for early differential diagnosis of PD in more rigorous studies.

There is still a long way to go before radiomic approaches can find widespread applications in clinical practice. This is mainly due to major deficiencies in the experimental design and pitfalls in the assessment methodology that limit the reliability and reproducibility of most ML methods (23,24). There is also a lack of consensus among imaging specialists and clinicians regarding different radiomic features and a

Page 4 of 5

Peng et al. Differential diagnosis of parkinsonism with radiomics

wide array of classification algorithms. This issue may be resolved by directly comparing different methods and using the same outcome measures in large samples of parkinsonian patients whose clinical diagnoses are uncertain initially but confirmed subsequently. To ensure the total impartiality of the outcomes in such endeavors it is essential to maintain a double-blind study between imaging analysts and clinicians. It is important to bear in mind that the maximum accuracy that can be achieved by any means of diagnosis will be limited given the heterogeneity of PD or each major form of APD due to clinical subtypes, genetic variants and other coexisting conditions (25). More concerted efforts are necessary on standardization, optimization and approval of easily accessible ML diagnostic tools in conjunction with aligned professional societies, government regulators and commercial vendors. This process can be accelerated by sharing databases and analytical tools in public domain, with the ultimate goal of achieving excellent generalizability of the techniques and a high degree of agreement among stake holders.

Going forwards there is an urgent need to develop standardized radiomics methodology and diagnostic criteria based on prospective national and international clinical trials in real world patients suspected to have any form of parkinsonian disorders, with longitudinal evaluation of eventual clinical outcomes and preferably with pathologic confirmation (1). It would be of greater value to improve diagnostic accuracy by combining unique features corresponding to region- or network-based imaging biomarkers with FDG PET. There is also general consensus that accuracy for early and differential diagnosis can be further enhanced by including any additional demographic and clinical information available. Ultimately, such collective endeavors in the field will establish the true accuracy and error bounds of differential diagnosis in light of inherent and sometimes large variability in both biomarkers and clinical examination. The continued improvement in the performance of advanced ML diagnostic tools would promote the delivery of personalized medicine in parkinsonism and optimize the subject selection in clinical trials of targeted therapies.

Acknowledgments

Footnote

References

1. Tang CC, Poston KL, Eckert T, et al. Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis. Lancet Neurol 2010;9:149-58.
2. Niethammer M, Tang CC, Feigin A, et al. A disease-specific metabolic brain network associated with corticobasal degeneration. Brain 2014;137:3036-46.
3. Meyer PT, Frings L, Rucker G, et al. (18)F-FDG PET in Parkinsonism: Differential Diagnosis and Evaluation of Cognitive Impairment. J Nucl Med 2017;58:1888-98.
4. Woo CW, Chang LJ, Lindquist MA, et al. Building better biomarkers: brain models in translational neuroimaging. Nat Neurosci 2017;20:365-77.
5. Schindlbeck KA, Eidelberg D. Network imaging biomarkers: insights and clinical applications in Parkinson's disease. Lancet Neurol 2018;17:629-40.
6. Wu Y, Jiang JH, Chen L, et al. Use of radiomic features and support vector machine to distinguish Parkinson's disease cases from normal controls. Ann Transl Med 2019;7:773.
7. Wenzel M, Milletari F, Kruger J, et al. Automatic classification of dopamine transporter SPECT: deep convolutional neural networks can be trained to be robust with respect to variable image characteristics. Eur J Nucl Med Mol Imaging 2019;46:2800-11.

8. Habeck C, Foster NL, Perneczky R, et al. Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. Neuroimage 2008;40:1503-15.

9. Peng S, Eidelberg D, Ma Y. Brain network markers of abnormal cerebral glucose metabolism and blood flow in Parkinson's disease. Neurosci Bull 2014;30:823-37.

10. Peng S, Ma Y, Spetsieris PG, et al. Characterization of disease-related covariance topographies with SSMPCA toolbox: effects of spatial normalization and PET scanners. Hum Brain Mapp 2014;35:1801-14.

11. Meles SK, Renken RJ, Pagani M, et al. Abnormal pattern of brain glucose metabolism in Parkinson's disease: replication in three European cohorts. Eur J Nucl Med Mol Imaging 2020;47:437-50.

12. Matthews DC, Lerman H, Lukic A, et al. FDG PET Parkinson's disease-related pattern as a biomarker for clinical trials in early stage disease. Neuroimage Clin 2018;20:572-9.

13. Wu P, Wang J, Peng S, et al. Metabolic brain network in the Chinese patients with Parkinson's disease based on 18F-FDG PET imaging. Parkinsonism Relat Disord 2013;19:622-7.

14. Holtbernd F, Gagnon JF, Postuma RB, et al. Abnormal metabolic network activity in REM sleep behavior disorder. Neurology 2014;82:620-7.

15. Wu P, Yu H, Peng S, et al. Consistent abnormalities in metabolic network activity in idiopathic rapid eye movement sleep behaviour disorder. Brain 2014;137:3122-8.

16. Garraux G, Phillips C, Schrouff J, et al. Multiclass classification of FDG PET scans for the distinction between Parkinson's disease and atypical parkinsonian syndromes. Neuroimage Clin 2013;2:883-93.

17. Spetsieris PG, Ma Y, Dhawan V, et al. Differential diagnosis of parkinsonian syndromes using PCA-based functional imaging features. Neuroimage 2009;45:1241-52.

18. Spetsieris P, Ma Y, Peng S, et al. Identification of disease-related spatial covariance patterns using neuroimaging data. J Vis Exp 2013. doi: 10.3791/50319.

19. Tripathi M, Tang CC, Feigin A, et al. Automated Differential Diagnosis of Early Parkinsonism Using Metabolic Brain Networks: A Validation Study. J Nucl Med 2016;57:60-6.

20. LeWitt PA, Rezai AR, Leehey MA, et al. AAV2-GAD gene therapy for advanced Parkinson's disease: a double-blind, sham-surgery controlled, randomised trial. Lancet Neurol 2011;10:309-19.

21. Pagani M, Giuliani A, Oberg J, et al. Progressive Disintegration of Brain Networking from Normal Aging to Alzheimer Disease: Analysis of Independent Components of (18)F-FDG PET Data. J Nucl Med 2017;58:1132-9.

22. Chen K, Reiman EM, Huan Z, et al. Linking functional and structural brain images with multivariate network analyses: a novel application of the partial least square method. Neuroimage 2009;47:602-10.

23. Walker Z, Gandolfo F, Orini S, et al. Clinical utility of FDG PET in Parkinson's disease and atypical parkinsonism associated with dementia. Eur J Nucl Med Mol Imaging 2018;45:1534-45.

24. 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.

25. Williams DR, Litvan I. Parkinsonian syndromes. Continuum (Minneap Minn) 2013;19:1189-212.