

Radiomics-Based Machine Learning to Support Visual Assessment for Improved Epilepsy Classification Using ^{18}F -FDG Brain PET

Naghmeh Firouzi^{1,3}, Hossein Ghadiri^{2,3*}, Elnaz Jenabi⁴, Parham Geramifar⁴

1. Department of Medical Physics, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
2. Department of Medical Physics and Biomedical Engineering, Tehran University of Medical Sciences, Tehran, Iran
3. Research Center for Molecular and Cellular Imaging, Advanced Medical Technologies and Equipment Institute, Tehran University of Medical Sciences, Tehran, Iran
4. Research Center for Nuclear Medicine, Tehran University of Medical Sciences, Tehran, Iran

ARTICLE INFO	ABSTRACT
Article type: Original Paper	Introduction: This study explored radiomics-based machine learning (ML) models as complementary tools to visual evaluation for classifying drug-resistant epilepsy patients and healthy controls using ^{18}F -FDG brain Positron Emission Tomography (PET). Because visual interpretation can be subjective and variable, especially for novice readers, objective and reproducible computational methods are needed.
Article history: Received: Oct 30, 2024 Accepted: May 30, 2025	Material and Methods: Twenty-one drug-resistant epilepsy patients and sixteen healthy controls underwent ^{18}F -FDG brain PET imaging. From contralateral brain regions, 92 radiomics features (first-order statistics and second-order texture matrices) were extracted. Feature selection included Student's t-test, principal component analysis, and ridge regression. Logistic regression (LR) and support vector machine (SVM) classifiers were trained and evaluated using 10-fold cross-validation and repeated 80/20 train-test splits. A permutation test ($n = 1000$) assessed whether differences between classifier performances were statistically significant. LR, chosen for its lower computational cost and interpretability, was used for comparison with human visual assessments.
Keywords: Radiomics Brain PET Imaging Epilepsy Classification Machine Learning	Results: Across six radiomics feature groups, LR models demonstrated strong performance, with mean accuracy of 0.94(0.05), precision 0.96(0.03), recall 0.92(0.10), specificity 0.97(0.02), and AUC 0.98(0.00). SVM models showed similarly high accuracy 0.98(0.01), precision 0.94(0.05), recall 0.96(0.03), specificity 0.98(0.01), and AUC 0.98(0.00). Novice visual assessments had moderate accuracy (0.62 and 0.67), perfect specificity, lower sensitivity (0.60 and 0.65), and AUCs of 0.80 and 0.825. The final LR model achieved a mean AUC of 0.96(0.01). Conclusion: This hybrid radiomics-visual approach improves classification accuracy in pre-surgical evaluation of drug-resistant epilepsy. By integrating quantitative radiomics with clinical interpretation, the framework reduces variability and improves reliability for less experienced clinicians.

► Please cite this article as:

Firouzi N, Ghadiri H, Jenabi E, Geramifar P. Radiomics-Based Machine Learning to Support Visual Assessment for Improved Epilepsy Classification Using ^{18}F -FDG Brain PET. Iran J Med Phys 2025; 22: 218-229. 10.22038/ijmp.2025.83024.2461.

Introduction

Epilepsy, one of the most common neurological disorders, affects about 50 million people worldwide [1]. It is characterized by recurrent seizures and is linked to neurobiological, cognitive, and psychosocial impairments [2]. Its causes are classified as genetic, structural, metabolic, infectious, immune-mediated, or idiopathic [3]. About 30–40% of patients develop drug-resistant epilepsy (DRE), defined as persistent seizures despite two well-tolerated regimens [2, 4, 5]. Surgical resection of the epileptogenic zone (EZ) can provide seizure control, requiring accurate presurgical localization [6, 7]. Structural magnetic resonance imaging (MRI) can assist in identifying the underlying cause and localizing the epileptogenic zone. Neuropsychological testing, functional magnetic

resonance imaging (fMRI), ^{18}F -Fluorodeoxyglucose (^{18}F -FDG) imaging, tractography, magnetoencephalography (MEG), and a fusion of fMRI and electroencephalography (EEG) are among the assessments employed [8]. MRI detects epileptogenic lesions in up to half of DRE patients, though subtle lesions may be missed without expert review [9, 10]. Interictal ^{18}F -FDG PET reveals hypometabolism in epileptogenic regions and aids diagnosis when MRI or EEG results are inconclusive [8, 11]. Identifying the seizure onset zone (SOZ) is essential for surgical decision-making [11]. Despite its advantages, ^{18}F -FDG brain positron emission tomography (PET) is often interpreted qualitatively through visual analysis [12]. Interpretation accuracy depends on reader expertise

*Corresponding Author: Tel: +98-21-8898 9487; Fax: +98-216405 3431; E-mail: H-ghadiri@tums.ac.ir

and image quality, making complementary tools especially valuable for novice physicians [13].

Radiomics extracts quantitative features beyond human perception. Combined with machine learning (ML), it enables automated, objective pattern recognition increasingly applied in epilepsy imaging [14-18].

We hypothesized that observer expertise affects visual assessment and developed a radiomics-based ML framework to complement it.

Radiomics extracts high-dimensional data from medical images, capturing subtle metabolic patterns that are often missed by conventional visual evaluation. ML algorithms, such as logistic regression (LR), then transform these features into actionable diagnostic insights. Our framework integrates computational precision with clinical context, combining ML output and expert interpretation for objective yet relevant assessment. Previous studies have mostly focused on isolated approaches. For example, Liu et al. (2018) used radiomics to predict epilepsy in glioma patients but did not integrate clinical evaluations, limiting its practical use [19]. On the other hand, Avendaño-Estrada et al. relied only on visual interpretations of ^{18}F -FDG PET images, which can vary between observers [12]. Our work bridges this gap by using radiomics to augment, not replace, clinical expertise. By validating radiomics features against expert consensus (ground truth), we ensure that the computational models align with clinically meaningful patterns for classifying DRE patients from HCs.

We developed ML models using radiomics features from ^{18}F -FDG brain PET images to complement visual assessments and assist novice physicians in distinguishing DRE patients from HCs.

Materials and Methods

Patients' population and PET imaging parameters

In this retrospective study, we analyzed ^{18}F -FDG brain PET images from 21 patients with refractory epilepsy (mean age: 55 ± 11 years; 13 males [61.9%] and 8 females [38.1%]). The sample size was determined based on the availability of eligible imaging data and strict inclusion/exclusion criteria. As a

retrospective study, no prior power analysis was performed; however, robust cross-validation (CV) and feature selection techniques were applied to minimize overfitting and improve generalizability.

All patients had prior EEG/MRI evaluations that failed to localize the EZ and presented with temporal or extra-temporal epilepsy. Of 31 initial ^{18}F -FDG PET scans, 10 generalized epilepsy cases were excluded. Figure 1 shows the inclusion/exclusion flowchart. All scans were acquired on a Biograph PET/CT scanner (Siemens Healthcare, Erlangen, Germany) at Shariati Hospital, Tehran, Iran, following manufacturer-recommended calibration and acquisition protocols. Patients rested in a dimly lit room for 40 min before receiving ~ 370 MBq of ^{18}F -FDG. PET acquisition began 60 min post-injection as single-bed, 10-min scans, with low-dose computed tomography (CT) (80–130 kV, 50–80 mAs) for attenuation correction. Images were reconstructed using matrix sizes of $336 \times 336 \times 110$ with voxel sizes of $1.018 \times 1.018 \times 3$ mm³.

The control group included 16 healthy controls (HCs) (age = 56 ± 8.1 years; 11 females) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

ADNI, launched in 2003, integrates imaging, biomarkers, and cognitive testing to monitor mild cognitive impairment (MCI) and early Alzheimer's Disease (AD) progression (www.adni-info.org).

Based on ADNI's documentation of PET acquisition protocols, the HC category used in this study had been injected ^{18}F -FDG (mean 185 MBq) intravenously, and the images had been reconstructed in a $336 \times 336 \times 109$ matrix with the voxel size of $1.018 \times 1.018 \times 2.02$ mm³. Table 1 shows PET imaging parameters and acquisition settings.

The study received ethical approval from the Ethics Committee of Shahid Sadoughi University of Medical Sciences (Code: IR.SSU.MEDICINE.REC.1395.293) and complied with the revised Declaration of Helsinki. Participants provided informed consent. All personal identifiers were removed, and data were securely stored. ADNI data followed its standard anonymization protocols. Figure 1 outlines participant inclusion and exclusion.

Table 1. PET Imaging Parameters and Acquisition Settings Used in This Study

Parameter	Shariati Hospital (Patients)	ADNI Database (Healthy Controls)
PET Scanner Model	Siemens Biograph PET/CT	Siemens Biograph PET/CT
Injected Dose (MBq)	~ 370 MBq	~ 185 MBq
Pre-scan Rest Time	60 min	30 min
Scan Duration	10 min	30 min
Matrix Size	$336 \times 336 \times 110$	$336 \times 336 \times 109$
Voxel Size (mm ³)	$1.018 \times 1.018 \times 3$	$1.018 \times 1.018 \times 2.02$
Attenuation Correction	Low-dose CT	Low-dose CT
Reconstruction Parameters	Iterative: OSEM-3D	Iterative: OSEM-3D

OSEM-3D: Three-dimensional ordered subsets expectation-maximization. PET: Positron Emission Tomography. CT: computed tomography.

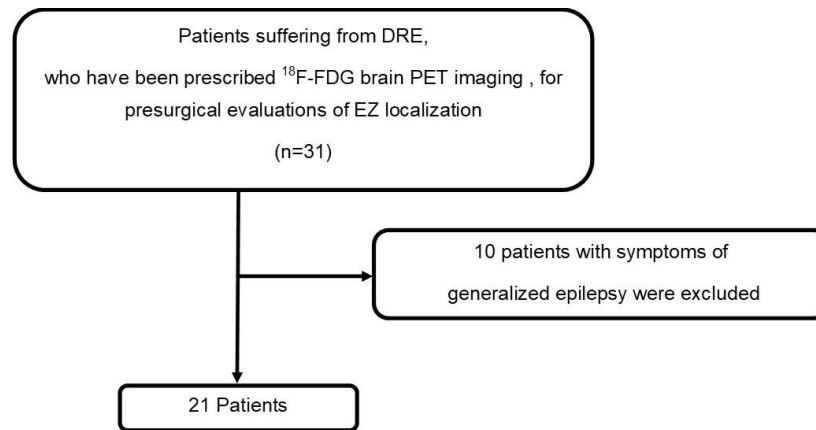


Figure 1. Flow diagram of subject inclusion and exclusion criteria. DRE: Drug-Resistant Epilepsy. ^{18}F -FDG: F18 Fludeoxyglucose. PET: Positron Emission Tomography

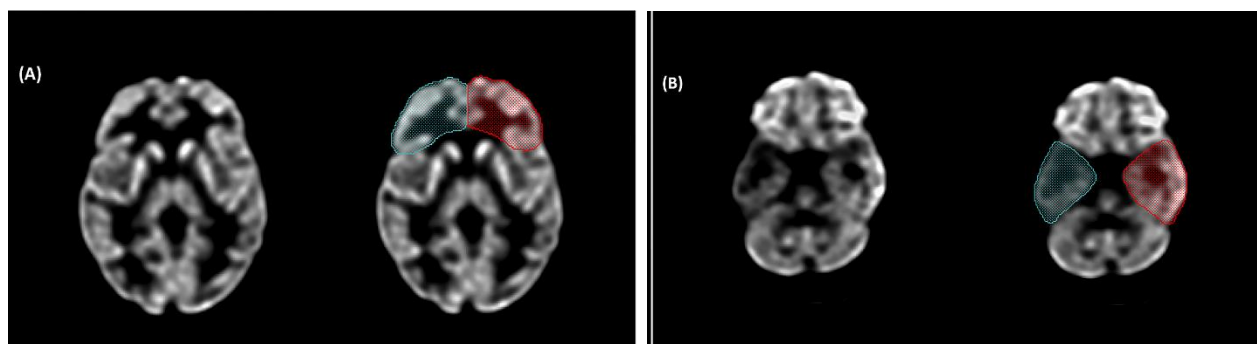


Figure 2. Illustrations of manually drawn regions of interest in contralateral brain lobes of a healthy control (A) and an epilepsy patient (B) through ^{18}F Fludeoxyglucose Positron Emission Tomography images.

Segmentation and Image Preprocessing

Before segmentation, a median filter (Iteration =3, XY planes) was applied for image smoothing and noise reduction. The drawn regions of interest (ROIs) in patients' ^{18}F -FDG brain PET images encompassed the suspected brain lobes, including SOZs, and their corresponding contralateral lobes. For HCs, the same procedure was applied to randomly selected contralateral lobes on both hemispheres. The segmentation was performed manually and meticulously, with each slice segmented individually using Avizo software (version 2019, RRID: SCR_014431). Figure 2 is an example of regions of interest in a healthy control and an epilepsy patient on ^{18}F -FDG PET images.

ROIs were delineated by consensus between two nuclear medicine (NM) physicians, manually segmented, and independently reviewed in a blinded manner by a senior NM specialist. Although inter-/intra-rater variability was not quantified, expert supervision ensured consistency. The thresholds used, including the 50% threshold, were also determined by consensus and applied uniformly across all cases.

The identification of the suspected lobes in patients to draw ROIs relied on the consensus of two expert NM physicians' interpretations, serving as the ground truth for the study. All assessments were blinded to patient data.

The masks and the ^{18}F -FDG brain PET images were saved in meta image header and raw data format (MHA) and used for radiomics feature extraction. Imaging protocols were standardized to minimize variability, including resampling all images to a uniform voxel size ($1 \times 1 \times 1 \text{ mm}^3$) before radiomics feature extraction. Inter-session variability was minimized by excluding scans with motion artifacts or incomplete coverage.

During the preprocessing phase, missing data points were not a concern as the dataset was derived from complete and validated ^{18}F -FDG brain PET imaging data. Outliers were handled by conducting thorough statistical inspections during feature selection, and features displaying significant inconsistencies were excluded. Z-score normalization was applied to standardize feature values, ensuring uniform scaling for ML model training. This process standardized each feature by centering it around zero and scaling it based on its standard deviation, which is critical for optimizing the performance of ML models, such as LR and support vector machine (SVM).

Radiomics feature extraction

92 features were extracted from each ROI on ^{18}F -FDG brain PET images using an open-source package (PyRadiomics, v3.0.1): [1] 19 first-order (F-order) features, and (2) 73 second-order features not considering the shape features (including gray level co-

occurrence matrix (GLCM), gray level run-length matrix (GLRLM), gray-level size zone matrix (GLSZM), neighboring gray tone difference matrix (NGTDM), and gray level dependence matrix (GLDM)). Notably, most features in the PyRadiomics package adhere to the feature definitions outlined by the Imaging Biomarker Standardization Initiative (IBSI) [20].

Features were extracted from 878 patient slices (affected/unaffected lobes) and 884 HC slices (bilateral random lobes).

Feature selection and normalization

Feature selection was conducted to identify the most informative and non-redundant radiomics features. First, a statistical filtering process (e.g., Student's t-test) was used to determine whether features showed significant differences between contralateral ROIs in patients and HCs. A total of 92 radiomics features were extracted, including F-order and second-order texture feature. The small dataset (21 patients, 16 HCs) yielded a ~2.5:1 feature-to-sample ratio, increasing the risk of overfitting.

To address this, dimensionality reduction was necessary to improve model generalizability. First, univariate feature selection (Student's t-test) removed non-discriminative features, reducing redundancy. Then, Principal component analysis (PCA) was applied separately to each feature group, retaining 95% of the variance while mitigating multicollinearity. This threshold is commonly used in radiomics and neuroimaging studies to balance dimensionality reduction with information retention. PCA reduced redundancy by transforming features into uncorrelated components [21, 22], followed by ridge regression (L2 regularization) to enhance model robustness.

Ridge regression adds a penalty term (L2 norm) to the loss function, which shrinks the coefficients of correlated features toward zero. This prevents any single feature from dominating the model and ensures stability and generalizability, particularly in high-dimensional datasets. Together, PCA and ridge regression enhanced interpretability and retained key diagnostic information. After dimensionality reduction, 11 optimal features were retained: 2 GLCM, 2 GLDM, 1 GLRLM, 2 GLSZM, 1 NGTDM, and 3 F-order features. The final number of principal components retained for each feature group is as follows: 6 (F-order), 5 (GLCM), 5 (GLDM), 3 (GLRLM), 4 (GLSZM), 3 (NGTDM) (Table 4).

To ensure consistency in scale, z-score normalization was applied to all remaining features. Outlier handling was inherently addressed through feature selection techniques, and no imputation was needed as there were no missing values in the dataset.

Machine learning model building

In this study, LR and SVM were selected for their interpretability and suitability for small, high-dimensional dataset [23, 24]. Default hyperparameters

(SVM with linear kernel, LR with L2 regularization) were used to avoid overfitting. Alternative models were preliminarily tested but showed similar or lower accuracy, supporting our choice. Also, models such as Convolutional Neural Networks (CNNs) were not tested in this study. Deep models need larger datasets to generalize, unsuitable for our limited sample. Additionally, Gated Recurrent Units are primarily designed for sequential data, whereas our study focused on static imaging-derived radiomics features.

Stratified 10-fold cross-validation (CV) ensured balanced class representation and robust generalization. Each model was trained on nine folds and tested on one, with metrics averaged over 10 repetitions. The 10-fold approach offered a good balance between bias and variance for our limited sample [25], and performance was evaluated separately for each feature group.

To statistically evaluate whether the performance differences between LR and SVM were significant, permutation tests ($n = 1000$) were performed using the AUC metric across all radiomics feature groups. AUC was selected as the representative metric due to its robustness to class imbalance and widespread use in classifier evaluation. The methodology followed standard random label shuffling to generate a null distribution of performance differences.

LR was preferred for interpretability, linking coefficients directly to feature importance, which is critical for clinical translation. Additionally, LR is resource-efficient during deployment, making it more practical for real-world applications, and is widely adopted in medical studies due to its simplicity and reliability.

Subsequently, for LR implementation, the dataset was randomly divided into 80% training and 20% test sets. The training set was further split 80–20 for validation. The training set was used for model learning, while the validation set provided an estimate of model performance. Splitting was repeated 10 times to reduce variability, with fixed randomization. The feature selection process and classification methods were implemented in Python 3.7.0 (RRID: SCR_008394) utilizing the Scikit-Learn library (version 0.24.1).

Visual Assessment

Two novice NM physicians evaluated ^{18}F -FDG brain PET images slice by slice, independently. They assessed glucose metabolism asymmetry and uptake patterns in specific brain regions as well as the classification of DRE patients and HCs. The ground truth classification of patients and HCs was determined based on the consensus visual assessments of two expert NM physicians. Assessments were blinded; physicians had no access to patient data or prior EEG/MRI results. Performance metrics (accuracy, precision, recall, specificity, AUC) were calculated for visual assessments.

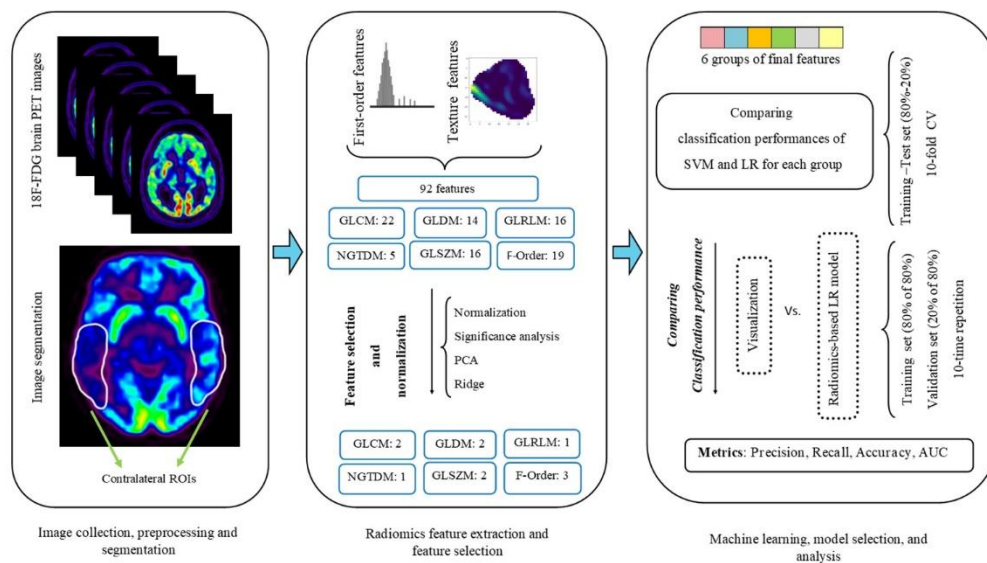


Figure 3. Flow diagram of radiomics models construction procedure. PET: Positron Emission Tomography. ROIs: Regions of Interest. GLCM: Gray Level Co-Occurrence Matrix. GLDM: Gray Level Dependence Matrix .GLRLM: Gray Level Run-Length Matrix .NGTDM: Neighboring Gray Tone Difference Matrix.GLSZM: Gray-Level Size Zone Matrix .F-Order: First-Order. PCA: Principal Component Analysis. SVM: Support Vector Machine. LR: Logistic Regression. CV: Cross Validation. AUC: Area Under the Curve.

Performance evaluations and Statistical analysis

Model performance was assessed via accuracy, precision, recall, specificity, and AUC for each fold; mean and standard deviation (SD) across folds quantified overall LR and SVM performance.

For the final LR model, metrics were calculated across all feature groups and compared with AUCs from novice physicians' visual assessments. Figure 3 illustrates the step-by-step process of constructing our radiomics-based ML models.

Mean $\pm$ SD and group differences in age and sex between patients and HCs were analyzed. The Student's t-test was used for univariate analysis of baseline characteristics and continuous variables. The Fisher exact test was employed to evaluate the statistical differences in categorical variables. Statistical analysis was conducted using SPSS software (version 26.0; IBM, RRID: SCR_019096), with significance set at $p < 0.05$.

Results

Study Participants

Table 2 summarizes participant demographics. The mean ages of patients and HCs were 55 and 56 years, with no significant group differences in age or sex.

P-values were achieved by a two-tailed t-test (age) and Fisher exact test (gender). *: Number. SD (Standard Deviation).

Table 3 presents performance metrics from visual assessments by two novice NM physicians (Visualization1 and Visualization2) for classifying DRE patients and HCs.

Table 4 lists excluded and retained features for each group. Features showing significant asymmetry in HCs and non-significant differences in patients were removed. The ultimate features were determined by applying PCA followed by ridge regression on each group of features.

Table 1. Demographic characteristics of patients and healthy controls

Characteristics	Patients	Healthy Controls	P-value
Numbers (%)	21 (56.7)	16(43.2)	
Gender			
#* of male	13	5	0.09
# of female	8	11	
Age (mean (SD))	55 (11)	56(8.1)	>0.53

Table 3. Calculated accuracy, precision, sensitivity, specificity, and area under the curve (AUC) metrics for classification of drug-resistant epilepsy patients and healthy controls based on visualization of two novice nuclear medicine physicians

Reader	Accuracy	Precision	Recall (Sensitivity)	Specificity	AUC
Visualization1	.62	1	0.6	1	0.8
Visualization2	.67	1	0.65	1	0.825

Table 4. Names and numbers of the excluded and retained features

Excluded features	Feature categories		Features' names	Number
Provided Significant differences in contralateral ROIs in HCs' group.	F-order		10Percentile, Kurtosis, Interquartile Range, Entropy, Minimum, Robust Mean Absolute Deviation	6
	GLCM		Correlation, Joint Energy, Joint Entropy, Inverse Difference Moment Normalized, Inverse Difference Normalized, Maximum Probability, Sum Entropy	7
	GLDM		Dependence Non-Uniformity, Dependence Entropy, Low Gray Level Emphasis, Large Dependence Low Gray Level Emphasis	4
	GLRLM		Long Run Low Gray Level Emphasis, Run Entropy, Run Length Non-Uniformity	3
	GLSZM		Gray Level Non-Uniformity Normalized, Large Area Low Gray Level Emphasis, Size Zone Non-Uniformity, Zone Entropy	4
	NGTDM		Strength	1
	Provided non-significant difference in contralateral ROIs in patients' group.	F-order		Minimum, 10Percentile
GLCM			Correlation, Inverse Difference Moment Normalized, Inverse Difference Normalized, Maximum Probability, Informational Measure of Correlation2	5
GLDM			Dependence Non-Uniformity, Low Gray Level Emphasis, Large Dependence Low Gray Level Emphasis, Small Dependence Low Gray Level Emphasis	4
GLRLM			Gray Level Non-Uniformity Normalized, Long Run Low Gray Level Emphasis, Low Gray Level Run Emphasis, Run Length Non-Uniformity, Short Run Low Gray Level Emphasis	5
GLSZM			Gray Level Non-Uniformity Normalized, Large Area Low Gray Level Emphasis, Low Gray Level Zone Emphasis, Size Zone Non-Uniformity, Small Area Low Gray Level Emphasis	5
NGTDM			Busyness	1
Retained features		Feature categories	Number of components for PCA retaining 95% variance	Features' names
Final selected features provided non-significant difference in contralateral ROIs in HCs' group and significant differences in contralateral ROIs in patients' group.	F-order	6	90Percentile, Energy, Maximum	3
	GLCM	5	Autocorrelation, Joint Average	2
	GLDM	5	Dependence Non-Uniformity Normalized, Gray Level Non-Uniformity	2
	GLRLM	3	Gray Level Non-Uniformity	1
	GLSZM	4	Gray Level Variance, Large Area High Gray Level Emphasis	2
	NGTDM	3	Coarseness	1

F-order (First Order), GLCM (Gray Level Co-Occurrence Matrix), GLDM (Gray Level Dependence Matrix), GLRLM (Gray Level Run-Length Matrix), GLSZM (Gray-Level Size Zone Matrix), HCs (Healthy Controls), NGTDM (Neighboring Gray Tone Difference Matrix), PCA (principal component analysis), ROIs (regions of interest)

Table 5 and Figure 4 summarize 10-fold CV results for LR and SVM across all feature groups. Both models showed comparable performance, with AUC differences ranging from -0.004 to 0.000 ($p = 0.92-1.00$). Given this similarity, LR was selected for its interpretability and simplicity.

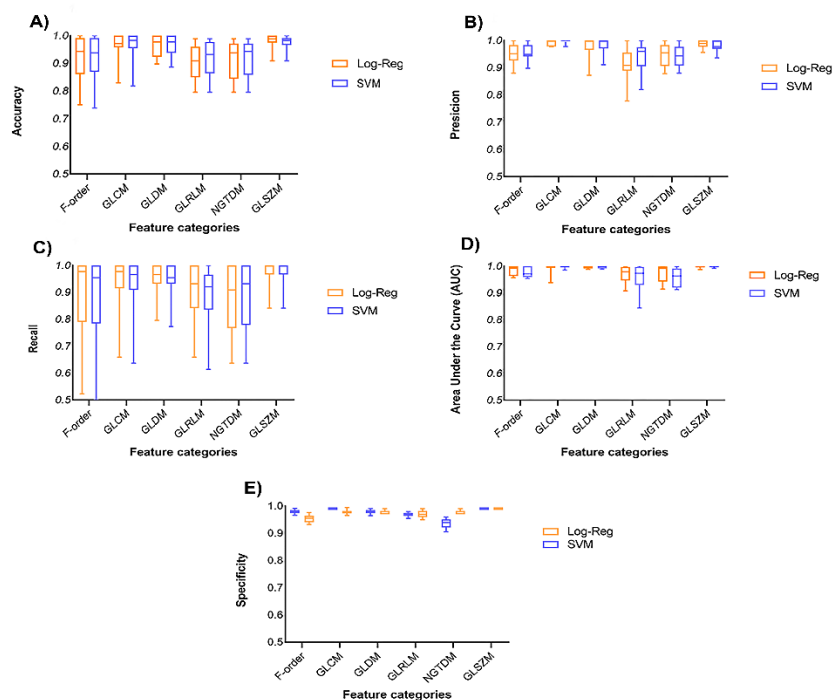


Figure 4. Boxplots of test dataset's Accuracy (A), Precision (B), Recall (C), Area under the curve (AUC) (D), and Specificity (E) including logistic regression (LR) and support vector machine (SVM) classifiers through 10-fold cross validation. GLCM: Gray Level Co-Occurrence Matrix. GLDM: Gray Level Dependence Matrix. GLRLM: Gray Level Run-Length Matrix. NGTDM: Neighboring Gray Tone Difference Matrix. GLSZM: Gray-Level Size Zone Matrix.

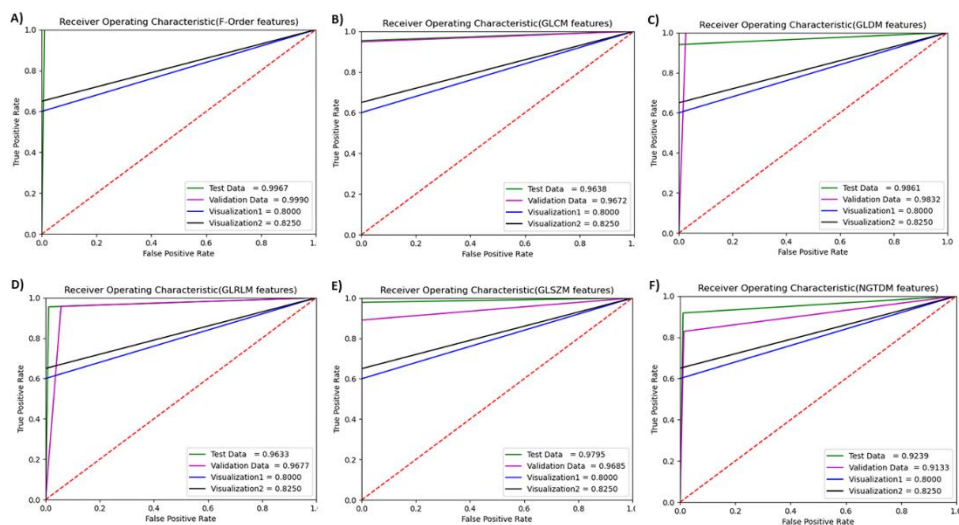


Figure 5. Mean area under the curve of the receiver operating characteristic (AUC-ROC) curves (with 10 times repetition) achieved by the logistic regression (LR) algorithm for the train and validation datasets together with the area under the curve of the receiver operating characteristic (AUC-ROC) curves of visual evaluations by two novice nuclear medicine physicians termed Visualization1 and Visualization2. Mean AUC-ROC curves of the LR models included groups of first-order (F-order) features (A), Gray Level Co-Occurrence Matrix (GLCM) features (B), Gray Level Dependence Matrix (GLDM) features (C), Gray Level Run-Length Matrix (GLRLM) features (D), Gray-Level Size Zone Matrix (GLSZM) features (E), Neighboring Gray Tone Difference Matrix (NGTDM) features (F).

Table 6 presents LR performance on validation and test sets (80–20 train–test, 80–20 validation split), reporting mean $\pm$ SD for all metrics.

Tables 5 and 6 present results using different evaluation strategies. Table 5 reports the performance of both LR and SVM models based on 10-fold CV, providing a robust estimate of generalizability. Table 6 presents the

performance of the final LR classifier on validation and test sets of each group using an 80–20 train–test split with 80–20 training–validation subsets.

Figure 5 displays AUC- Receiver Operating Characteristic (ROC) curves for LR models (test and validation sets) alongside novice NM physicians' visualizations.

Table 5. Classification (healthy controls, patients) performance indices using support vector machine (SVM) and logistic regression (LR) with 10-fold cross validation across feature categories. Mean (Standard Deviation) values are shown.

Datasets	Classifiers	Feature categories	Accuracy	Precision	Recall (Sensitivity)	Specificity	AUC
Train	SVM	F-order	0.9 ^v (0.00)	0.9 [^] (0.00)	0.9 [^] (0.00)	0.98(0.00)	0.99(0.00)
		GLCM	0.97(0.00)	0.99(0.00)	0.94(0.00)	0.99(0.00)	0.99(0.00)
		GLDM	0.96(0.00)	0.98(0.00)	0.95(0.00)	0.98(0.00)	0.99(0.00)
		GLRLM	0.91(0.00)	0.93(0.00)	0.88(0.01)	0.97(0.00)	0.95(0.00)
		GLSZM	0.97(0.00)	0.98(0.00)	0.95(0.00)	0.99(0.00)	0.99(0.00)
		NGTDM	0.91(0.00)	0.94(0.00)	0.89(0.01)	0.94(0.00)	0.96(0.00)
	Test	F-order	0.91(0.08)	0.95(0.03)	0.87(0.16)	0.98(0.01)	0.97(0.01)
		GLCM	0.96(0.05)	0.99(0.00)	0.92(0.11)	0.99(0.00)	0.99(0.00)
		GLDM	0.96(0.03)	0.98(0.02)	0.94(0.06)	0.98(0.01)	0.99(0.00)
		GLRLM	0.91(0.05)	0.94(0.05)	0.89(0.11)	0.97(0.01)	0.95(0.04)
		GLSZM	0.97(0.02)	0.97(0.02)	0.97(0.05)	0.99(0.00)	0.99(0.00)
		NGTDM	0.91(0.06)	0.94(0.04)	0.88(0.13)	0.94(0.03)	0.95(0.03)
Train	LR	F-order	0.9 [^] (0.00)	0.9 ^v (0.00)	0.9 [^] (0.00)	0.97(0.00)	0.99(0.00)
		GLCM	0.96(0.00)	0.98(0.00)	0.94(0.00)	0.98(0.00)	0.99(0.00)
		GLDM	0.96(0.00)	0.98(0.00)	0.95(0.00)	0.98(0.00)	0.99(0.00)
		GLRLM	0.90(0.00)	0.90(0.00)	0.91(0.00)	0.97(0.00)	0.96(0.00)
		GLSZM	0.98(0.00)	0.98(0.00)	0.97(0.00)	0.99(0.00)	0.99(0.00)
		NGTDM	0.91(0.00)	0.94(0.00)	0.87(0.01)	0.94(0.00)	0.97(0.00)
	Test	F-order	0.92(0.08)	0.94(0.03)	0.88(0.15)	0.97(0.02)	0.98(0.01)
		GLCM	0.96(0.05)	0.99(0.01)	0.93(0.10)	0.98(0.01)	0.99(0.01)
		GLDM	0.96(0.04)	0.97(0.04)	0.95(0.06)	0.98(0.01)	0.99(0.00)
		GLRLM	0.90(0.06)	0.91(0.06)	0.90(0.10)	0.97(0.02)	0.96(0.03)
		GLSZM	0.98(0.02)	0.98(0.01)	0.97(0.05)	0.99(0.00)	0.99(0.00)
		NGTDM	0.91(0.06)	0.94(0.04)	0.87(0.13)	0.94(0.03)	0.97(0.03)

AUC (Area under the curve), F-order (First Order), GLCM (Gray Level Co-Occurrence Matrix), GLDM (Gray Level Dependence Matrix), GLRLM (Gray Level Run-Length Matrix), GLSZM (Gray-Level Size Zone Matrix), HCs (Healthy Controls), NGTDM (Neighboring Gray Tone Difference Matrix)

Table 6. Classification (healthy controls, patients) performance indices by logistic regression (LR) for each group of features. Mean (Standard Deviation) values are shown

Datasets	Classifier	Feature categories	Accuracy	Precision	Recall (Sensitivity)	Specificity	AUC
Test	LR	F-order	0.95(0.01)	0.96(0.02)	0.94(0.03)	0.99(0.00)	0.99(0.00)
		GLCM	0.96(0.01)	0.98(0.01)	0.94(0.02)	0.98(0.00)	0.96(0.01)
		GLDM	0.96(0.01)	0.9 ^v (0.01)	0.95(0.02)	0.99(0.00)	0.98(0.01)
		GLRLM	0.89(0.02)	0.90(0.02)	0.89(0.04)	0.95(0.01)	0.96(0.02)
		GLSZM	0.97(0.01)	0.98(0.01)	0.97(0.01)	1(0.00)	0.97(0.01)
		NGTDM	0.90(0.01)	0.93(0.03)	0.86(0.03)	0.97(0.01)	0.92(0.01)
Validation		F-order	0.96(0.01)	0.96(0.02)	0.95(0.01)	0.98(0.01)	0.99(0.01)
		GLCM	0.96(0.01)	0.99(0.00)	0.93(0.02)	0.99(0.01)	0.96(0.01)
		GLDM	0.9 [^] (0.01)	0.9 ^v (0.01)	0.95(0.02)	0.97(0.01)	0.98(0.01)
		GLRLM	0.90(0.02)	0.91(0.03)	0.89(0.03)	0.91(0.02)	0.96(0.02)
		GLSZM	0.98(0.00)	0.98(0.00)	0.98(0.01)	0.99(0.00)	0.96(0.00)
		NGTDM	0.92(0.02)	0.96(0.01)	0.87(0.04)	0.95(0.01)	0.91(0.01)

AUC (Area under the curve), F-order (First Order), GLCM (Gray Level Co-Occurrence Matrix), GLDM (Gray Level Dependence Matrix), GLRLM (Gray Level Run-Length Matrix), GLSZM (Gray-Level Size Zone Matrix), HCs (Healthy Controls), NGTDM (Neighboring Gray Tone Difference Matrix)

Discussion

Epilepsy affects about 1% of the global population, and identifying the EZ accurately is critical for successful surgical outcomes in DRE [26-28]. ¹⁸F-FDG PET plays a key role in presurgical evaluation by revealing metabolic abnormalities that correlate with seizure onset zones and surgical prognosis [29-31]. However, conventional visual assessment of PET

images is often subjective and highly dependent on physician expertise, making subtle abnormalities difficult to detect [13, 32-34]. Recent advances in AI and ML have shown great promise in enhancing diagnostic precision, reducing observer variability, and supporting novice physicians in clinical decision-making [23, 35-38].

We applied radiomics analysis of ^{18}F -FDG brain PET images to classify DRE patients and HCs. To identify the most informative features, we excluded those with significant differences between contralateral ROIs in HCs ($p < 0.05$) and non-significant differences in patients ($p > 0.05$), followed by principal component analysis (PCA) and ridge regression to minimize redundancy and multicollinearity (Table 4).

SVM and LR were evaluated using stratified 10-fold CV for each feature group. Given the exploratory nature of this work, confidence intervals and hypothesis testing were omitted to emphasize model feasibility and performance metrics calculated through robust 10-fold CV [25].

As shown in Table 5, both LR and SVM achieved high and comparable performance across all feature groups (AUC = 0.95–0.99), confirming their strong discriminative ability.

Table 3 shows visual assessments by two novice NM physicians achieved moderate accuracy (0.62–0.67, AUC = 0.80–0.83) due to limited recall. In contrast, radiomics-based ML models (LR and SVM) reached mean AUC = 0.98 ± 0.00 (Table 5), confirming their superior reliability and value for less-experienced clinicians.

Permutation testing showed no significant AUC differences between LR and SVM (Table 5 and Figure 4), supporting the choice of LR as the final model for its interpretability and efficiency. The dataset was split into 80–20 training–testing subsets, and 80–20 training–validation subsets, repeated 10 times.

Table 6 represents promising performances for the test dataset. These findings suggest that the choice of radiomics feature group can substantially influence model performance. F-order features achieved the highest AUC (0.99 ± 0.00), indicating strong discriminative power for distinguishing patients with epilepsy from HCs, outperforming features like NGTDM. Additionally, GLSZM features effectively captured spatial heterogeneity in glucose uptake, with metrics such as Large Area High Gray Level Emphasis contributing to the classification of epileptogenic regions. Although interictal PET imaging typically reveals hypometabolism, this feature may reflect subtle regional asymmetries or relatively preserved glucose intensities that were informative for distinguishing DRE patients from HCs.

The AUC values for the classification of patients and HCs based on the visualizations implemented by the two novice NM physicians, referred to as Visualization1 and Visualization2 (shown in Figure 5), were 0.8 and 0.825, respectively. Radiomics-based ML models significantly outperformed novice physicians (AUC = 0.98 vs. 0.81), demonstrating superior reliability and sensitivity for DRE classification.

Among feature groups, NGTDM had the lowest AUC = 0.92 ± 0.01 , and F-order the highest AUC = 0.99 ± 0.00 , both outperforming novice visualizations (mean AUC = 0.81). Thus, for classifying DRE and HCs, besides physicians' visual interpretations, it can be

helpful to define suspected contralateral brain lobes and implement such radiomics-based LR models to boost the diagnosis accuracy.

This study demonstrated that radiomics-based ML models significantly outperformed novice visual assessments in classifying DRE patients and HCs from ^{18}F -FDG brain PET images.

The LR model achieved a mean $\pm$ SD AUC of 0.96 ± 0.01 , outperforming novice NM physicians (AUC: 0.80–0.825) (Table 3, Figure 5). Despite achieving high specificity and precision in the visual assessments, their moderate accuracy highlights limitations in identifying all affected patients. This underscores the potential value of radiomics models as adjunct tools, offering consistent sensitivity while maintaining interpretability. This gap highlights radiomics' ability to detect subtle metabolic asymmetries (e.g., in [GLSZM] features) that novices often overlook, directly addressing the study's goal of reducing diagnostic variability.

Mean values of precision (0.95 ± 0.01) and recall (0.95 ± 0.01) among the six feature categories indicate the model minimizes both false positives (misclassifying HCs) and false negatives (missing DRE cases), critical for avoiding unnecessary invasive procedures or delayed surgeries.

In addition to accuracy, precision, recall, and AUC, we emphasized specificity for its key role in correctly identifying HCs and minimizing false positives. Including this metric provided a more balanced evaluation of model performance, enhancing interpretability and clinical relevance in distinguishing DRE patients from HCs.

Radiomics-based ML models complement visual assessments by offering quantitative, objective insights that improve diagnostic accuracy and confidence—especially for novice physicians. They also accelerate image analysis, reduce observer variability, and streamline decision-making in presurgical evaluations.

Integrating radiomics-based ML into clinical workflows remains challenging due to inter-scanner variability, need for standardized preprocessing, and limited computational or technical resources in some settings. The "black-box" nature of these models can make it difficult for clinicians to trust and adopt them without clear explanations or visual aids.

Obtaining regulatory approval for ML tools in healthcare can be a long and resource-intensive process, requiring thorough validation. Clinicians also need training to use the tool effectively, and there may be resistance to new technologies. Further validation across diverse patient demographics and imaging centers is necessary to ensure the model's generalizability and reduce biases. Seamless integration with existing clinical systems is needed to avoid disrupting current practices. Addressing data privacy, security, and accountability in case of diagnostic errors is essential for compliance with ethical and legal standards.

Our high-performing ML models (LR, SVM) align with Hao et al. [39] and Liao et al. [40], though prior work mainly used MRI rather than PET-based

biomarkers. Compared to these works, our study provides novel evidence that PET-based radiomics can achieve superior classification performance. Hao et al. focused on localizing the epileptogenic zone using radiomics features from ^{18}F -FDG PET images [39]. Their model achieved a mean $\pm$ SD accuracy of 0.84 ± 0.03 , sensitivity of 0.78 ± 0.13 , specificity of 0.84 ± 0.04 , and a mean AUC of 0.89 ± 0.05 in identifying the epileptogenic zone in TLE patients. In the study by Liao et al. a radiomics-based ML model utilizing ^{18}F -FDG PET features achieved an accuracy of 0.948 sensitivity of 0.941 precision of 0.985, and an AUC of 0.984 in distinguishing temporal lobe epilepsy (TLE) patients from healthy controls [40]. Above mentioned studies achieved a high AUC for distinguishing between patients with temporal lobe epilepsy and healthy individuals using a radiomics-based ML model based on ^{18}F -FDG brain PET images. To review some studies exploring the use of radiomics in epilepsy as a central nervous system disease, in 2018, Liu et al. successfully developed and validated a prediction model for accurately predicting epilepsy incidence in low-grade glioma patients. The model demonstrated an accuracy of 0.793 in the primary cohort and 0.75 in the validation cohort, with an AUC of 0.875 for the primary cohort and 0.816 for the validation cohort.[19]. In 2019, Mo et al. presented a computational model that utilizes high-throughput radiomics features for the detection of MRI-negative hippocampal sclerosis as a histopathological hallmark and major underlying cause of temporal lobe epilepsy. They achieved an accuracy of 0.958, sensitivity of 0.963, specificity of 0.956, and AUCs of 0.997 in the primary cohort and 0.978 in the validation cohort [16]. In 2019, Zhuang et al. investigated the predictors of epilepsy presentation in unruptured brain arteriovenous malformations (bAVMs) through quantitative evaluation of location and radiomics features on T2-weighted imaging. They found that epilepsy-susceptible bAVMs exhibited distinct locations and radiomics features on T2-weighted imaging. The predictive score demonstrated an accuracy of 0.822 in the training dataset and 0.778 in the test dataset. The AUC was 0.866 with a 95% confidence interval of 0.791–0.940. On the training dataset, the sensitivity was 0.800 and the specificity was 0.850. On the test dataset, the sensitivity was 0.786, the specificity was 0.769 [41]. In 2021, Cheong et al. discovered significant differences in radiomics features between the affected and unaffected sides of patients with temporal lobe epilepsy (TLE). They found that radiomics analysis not only revealed abnormalities in the extrahippocampal regions of the affected side in TLE patients but also had the potential to identify MRI-negative cases of TLE. The extrahippocampal model achieved an AUC of 0.97 (95% CI, 0.94–1.00) on the training set, 0.90 (95% CI, 0.86–0.93) on the internal validation set, and 0.92 (95% CI, 0.86–0.98) on the external validation set. For the external validation, it showed a sensitivity of 0.92 and a specificity of 0.96 [42]. In a study in 2021, Gao et al. explored the relationship between radiomics features

and frontal glioma-associated epilepsy (GAE) and developed a robust radiomics-based model for predicting frontal GAE. Their findings indicate that radiomics analysis can effectively predict GAE without invasive procedures, enabling more accurate treatment planning for frontal glioma. Their proposed clinical-radiomics model holds promise for the precise prediction of frontal GAE. They achieved an accuracy of 0.82 sensitivity of 0.803 specificity of 0.840, and AUCs of 0.886 (0.819–0.940) in the training cohort and an accuracy of 0.782, sensitivity of 0.75, specificity of 0.815, and AUCs of 0.836 (0.707–0.937) in the test cohort [43].

Limitations and Future Work

This study has several limitations to consider. First, using PET data from two centers may have introduced variability due to differences in acquisition protocols and patient populations, even though both sites used the same scanner model. Although we applied standardized preprocessing steps, such as resampling, intensity normalization, and exclusion of asymmetric features in HCs, we did not formally harmonize inter-center variations. ComBat harmonization was not applied due to differing cohort distributions; future multi-site studies should assess and mitigate inter-center effects.

Second, all ROI delineations were determined based on expert consensus between two NM physicians to ensure consistency and reduce subjective variability. The segmentation was performed manually by an experienced professional and independently reviewed in a blinded manner by a senior NM specialist to validate accuracy and adherence to standardized anatomical guidelines. Although consistent, no formal rater-agreement metrics were computed; future work will include reproducibility analyses (e.g., Dice coefficient or ICC) and explore automated or semi-automated segmentation methods such as deep learning to improve standardization and reduce observer bias.

Third, although a combined-feature model using PCA was explored, it was excluded from the main analysis due to limited interpretability. Instead, we evaluated models based on individual feature groups to emphasize their distinct contributions. Future studies should consider advanced feature fusion techniques to enhance classification robustness while preserving explainability.

Finally, the small sample size limited subgroup analyses. Larger, multi-center studies are needed to validate findings and assess radiomics patterns across clinical subtypes.

Conclusion

In conclusion, the analysis reveals that ^{18}F -FDG brain PET radiomics-based ML models can serve as a complementary tool to enhance diagnostic accuracy in classifying patients and HCs during the pre-surgical evaluation of DRE patients. In addition to visual interpretations, when physicians delineate suspected contralateral brain lobes and apply the radiomics-based

ML models, it enhances diagnostic accuracy, especially benefiting less-experienced physicians. These radiomics-based ML models leverage quantitative imaging data to extract meaningful features, offering valuable insights that support more informed clinical decision-making. However, further validation and refinement of the models are essential to fully realize their potential and establish their clinical utility in enhancing the management of DRE.

Acknowledgment

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; 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 is 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 Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

The study received ethical approval from the ethics committee of Shahid Sadoughi University of Medical Sciences and Health Services (Code: IR.SSU.MEDICINE.REC.1395.293). The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

References

1. Rigney G, Lennon M, Holderrieth P. The use of computational models in the management and prognosis of refractory epilepsy: a critical evaluation. *Seizure*. 2021;91:132-40.
2. Milligan TA. Epilepsy: a clinical overview. *The American Journal of Medicine*. 2021;134(7):840-7.
3. Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: Position paper of the ILAE

- Commission for Classification and Terminology. *Epilepsia*. 2017;58(4):512-21.
4. Perucca E, Perucca P, White HS, Wirrell EC. Drug resistance in epilepsy. *The Lancet Neurology*. 2023;22(8):723-34.
5. Mohammadzadeh P, Nazarbaghi S. The prevalence of drug-resistant-epilepsy and its associated factors in patients with epilepsy. *Clinical neurology and neurosurgery*. 2022;213:107086.
6. Bernasconi N, Bernasconi A. Imaging the epileptic brain—time for new standards. *Nature Reviews Neurology*. 2014;10(3):133-4.
7. Kini LG, Thaker AA, Hadar PN, Shinohara RT, Brown M-G, Dubroff JG, Davis KA. Quantitative [18] FDG PET asymmetry features predict long-term seizure recurrence in refractory epilepsy. *Epilepsy & Behavior*. 2021;116:107714.
8. Thijs RD, Surges R, O'Brien TJ, Sander JW. Epilepsy in adults. *The Lancet*. 2019;393(10172):689-701.
9. Duncan JS, Winston GP, Koepp MJ, Ourselin S. Brain imaging in the assessment for epilepsy surgery. *The Lancet Neurology*. 2016;15(4):420-33.
10. Hainc N, McAndrews MP, Valiante T, Andrade DM, Wennberg R, Krings T. Imaging in medically refractory epilepsy at 3 Tesla: a 13-year tertiary adult epilepsy center experience. *Insights into Imaging*. 2022;13(1):99.
11. Steinbrenner M, Duncan JS, Dickson J, Rathore C, Wächter B, Aygun N, et al. Utility of 18F-fluorodeoxyglucose positron emission tomography in presurgical evaluation of patients with epilepsy: A multicenter study. *Epilepsia*. 2022;63(5):1238-52.
12. Avendaño-Estrada A, Velasco F, Velasco AL, Cuellar-Herrera M, Saucedo-Alvarado PE, Marquez-Franco R, et al. Quantitative analysis of [18F] FFMZ and [18F] FDG PET studies in the localization of seizure onset zone in drug-resistant temporal lobe epilepsy. *Stereotactic and Functional Neurosurgery*. 2019;97(4):232-40.
13. Wu D, Yang L, Gong G, Zheng Y, Jin C, Qi L, et al. Characterizing the hyper- and hypometabolism in temporal lobe epilepsy using multivariate machine learning. *Journal of Neuroscience Research*. 2021;99(11):3035-46.
14. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer*. 2012;48(4):441-6.
15. Fusco R, Granata V, Grazzini G, Pradella S, Borgheresi A, Bruno A, et al. Radiomics in medical imaging: pitfalls and challenges in clinical management. *Japanese journal of radiology*. 2022;40(9):919-29.
16. Mo J, Liu Z, Sun K, Ma Y, Hu W, Zhang C, et al. Automated detection of hippocampal sclerosis using clinically empirical and radiomics features. *Epilepsia*. 2019;60(12):2519-29.
17. Park JE, Kim HS. Radiomics as a quantitative imaging biomarker: practical considerations and the current standpoint in neuro-oncologic studies. *Nuclear medicine and molecular imaging*. 2018;52:99-108.
18. Sone D, Beheshti I. Clinical application of machine learning models for brain imaging in epilepsy: a review. *Frontiers in Neuroscience*. 2021;15:684825.

19. Liu Z, Wang Y, Liu X, Du Y, Tang Z, Wang K, et al. Radiomics analysis allows for precise prediction of epilepsy in patients with low-grade gliomas. *NeuroImage: Clinical*. 2018;19:271-8.
20. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJ, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328-38.
21. Salem N, Hussein S. Data dimensional reduction and principal components analysis. *Procedia Computer Science*. 2019;163:292-9.
22. Palo HK, Sahoo S, Subudhi AK. Dimensionality reduction techniques: Principles, benefits, and limitations. *Data Analytics in Bioinformatics: A Machine Learning Perspective*. 2021:77-107.
23. Segato A, Marzullo A, Calimeri F, De Momi E. Artificial intelligence for brain diseases: A systematic review. *APL Bioengineering*. 2020;4(4):041503.
24. Parmar C, Grossmann P, Bussink J, Lambin P. Machine learning methods for quantitative radiomic biomarkers. *Machine learning applications for Radiomics*. 2017;5:125.
25. Marcot BG, Hanea AM. What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Computational Statistics*. 2021;36(3):2009-31.
26. Guery D, Rheims S. Clinical management of drug resistant epilepsy: a review on current strategies. *Neuropsychiatric Disease and Treatment*. 2021:2229-42.
27. Sultana B, Panzini M-A, Carpentier AV, Comtois J, Rioux B, Gore G, et al. Incidence and prevalence of drug-resistant epilepsy: a systematic review and meta-analysis. *Neurology*. 2021;96(17):805-17.
28. Burkholder DB, Ritaccio AL, Shin C. Pre-surgical Evaluation. *Epilepsy*. 2021:345-65.
29. Fitsiori A, Hiremath SB, Boto J, Garibotto V, Vargas MI. Morphological and advanced imaging of epilepsy: beyond the basics. *Children*. 2019;6(3):43.
30. Guo J, Guo M, Liu R, Kong Y, Hu X, Yao L, et al. Seizure outcome after surgery for refractory epilepsy diagnosed by 18F-fluorodeoxyglucose positron emission tomography (18F-FDG PET/MRI): a systematic review and meta-analysis. *World Neurosurgery*. 2023;173:34-43.
31. Tomás J, Pittau F, Hammers A, Bouvard S, Picard F, Vargas MI, et al. The predictive value of hypometabolism in focal epilepsy: a prospective study in surgical candidates. *European journal of nuclear medicine and molecular imaging*. 2019;46:1806-16.
32. Pittau F, Grouiller F, Spinelli L, Seeck M, Michel CM, Vulliemoz S. The role of functional neuroimaging in pre-surgical epilepsy evaluation. *Frontiers in neurology*. 2014;5:31.
33. Peter J, Houshmand S, Werner TJ, Rubello D, Alavi A. Applications of global quantitative 18F-FDG-PET analysis in temporal lobe epilepsy. *Nuclear Medicine Communications*. 2016;37(3):223-30.
34. Wissel BD, Greiner HM, Glauser TA, Holland-Bouley KD, Mangano FT, Santel D, et al. Prospective validation of a machine learning model that uses provider notes to identify candidates for resective epilepsy surgery. *Epilepsia*. 2020;61(1):39-48.
35. Yuan J, Ran X, Liu K, Yao C, Yao Y, Wu H, Liu Q. Machine learning applications on neuroimaging for diagnosis and prognosis of epilepsy: A review. *Journal of neuroscience methods*. 2022;368:109441.
36. Cantor-Rivera D, Khan AR, Goubran M, Mirsattari SM, Peters TM. Detection of temporal lobe epilepsy using support vector machines in multi-parametric quantitative MR imaging. *Computerized Medical Imaging and Graphics*. 2015;41:14-28.
37. Memarian N, Kim S, Dewar S, Engel Jr J, Staba RJ. Multimodal data and machine learning for surgery outcome prediction in complicated cases of mesial temporal lobe epilepsy. *Computers in biology and medicine*. 2015;64:67-78.
38. Del Gaizo J, Mofrad N, Jensen JH, Clark D, Glenn R, Helpert J, Bonilha L. Using machine learning to classify temporal lobe epilepsy based on diffusion MRI. *Brain and behavior*. 2017;7(10):e00801.
39. Hao J, Xie Y, Liu Q, Xu J, Zhang P, editors. Localization of Epileptogenic Zone Based on Radiomics Features of 18F-FDG PET in Patients with Temporal Lobe Epilepsy. 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER); 2021: IEEE.
40. Liao K, Wu H, Jiang Y, Dong C, Zhou H, Wu B, et al. Machine learning techniques based on 18F-FDG PET radiomics features of temporal regions for the classification of temporal lobe epilepsy patients from healthy controls. *Frontiers in Neurology*. 2024;15:1377538.
41. Zhang Y, Yan P, Liang F, Ma C, Liang S, Jiang C. Predictors of epilepsy presentation in unruptured brain arteriovenous malformations: a quantitative evaluation of location and radiomics features on T2-weighted imaging. *World Neurosurgery*. 2019;125:e1008-e15.
42. Cheong E-N, Park JE, Jung DE, Shim WH. Extrahippocampal radiomics analysis can potentially identify laterality in patients with MRI-negative temporal lobe epilepsy. *Frontiers in neurology*. 2021;12:706576.
43. Gao A, Yang H, Wang Y, Zhao G, Wang C, Wang H, et al. Radiomics for the prediction of epilepsy in patients with frontal glioma. *Frontiers in oncology*. 2021;11:725926.