


Article

Improving the Classification of PCNSL and Brain Metastases by Developing a Machine Learning Model Based on ¹⁸F-FDG PET

Can Cui ¹, Xiaochen Yao ², Lei Xu ², Yuelin Chao ³, Yao Hu ¹, Shuang Zhao ¹, Yuxiao Hu ^{1,*} and Jia Zhang ¹

¹ Department of PET/CT Center, Jiangsu Cancer Hospital and Jiangsu Institute of Cancer Research and the Affiliated Cancer Hospital of Nanjing Medical University, Nanjing 210009, China

² Department of Nuclear Medicine, Nanjing First Hospital, Nanjing Medical University, Nanjing 210006, China

³ Department of Cardiology, Nanjing First Hospital, Nanjing Medical University, Nanjing 210006, China

* Correspondence: hyx_0102@yeah.net; Tel.: +86-2583284736

Abstract: Background: The characteristic magnetic resonance imaging (MRI) and the positron emission tomography (PET) findings of PCNSL often overlap with other intracranial tumors, making definitive diagnosis challenging. PCNSL typically shows iso-hypointense to grey matter on T2-weighted imaging. However, a particular part of PCNSL can demonstrate T2-weighted hyperintensity as other intracranial tumors. Moreover, normal high uptake of FDG in the basal ganglia, thalamus, and grey matter can mask underlying PCNSL in ¹⁸F-FDG PET. In order to promote the efficiency of diagnosis, the MRI-based or PET/CT-based radiomics models combining histograms with texture features in diagnosing glioma and brain metastases have been widely established. However, the diagnosing model for PCNSL has not been widely reported. The study was designed to investigate a machine-learning (ML) model based on multiple parameters of 2-deoxy-2-[¹⁸F]-fluoro-D-glucose (¹⁸F-FDG) PET for differential diagnosis of PCNSL and metastases in the brain. **Methods:** Patients who underwent an ¹⁸F-FDG PET scan with untreated PCNSL or metastases in the brain were included between May 2016 and May 2022. A total of 126 lesions from 51 patients (43 patients with untreated brain metastases and eight patients with untreated PCNSL), including 14 lesions of PCNSL, and 112 metastatic lesions in the brain, met the inclusion criteria. PCNSL or brain metastasis was confirmed after pathology or clinical history. Principal component analysis (PCA) was used to decompose the datasets. Logistic regression (LR), support vector machine (SVM), and random forest classification (RFC) models were trained by two different groups of datasets, the group of multi-class features and the group of density features, respectively. The model with the highest mean precision score was selected. The testing sets and original data were used to examine the efficacy of models separately by using the weighted average *F1 score* and area under the curve (AUC) of the receiver operating characteristic curve (ROC). **Results:** The multi-class features-based RFC and SVM models reached identical weighted-average *F1 scores* in the testing set, and the score was 0.98. The AUCs of RFC and SVM models calculated from the testing set were 1.00 equally. Evaluated by the original dataset, the RFC model based on multi-class features performs better than the SVM model, whose weighted-average *F1 scores* of the RFC model calculated from the original data were 0.85 with an AUC of 0.93. **Conclusions:** The ML based on multi-class features of ¹⁸F-FDG PET exhibited the potential to distinguish PCNSL from brain metastases. The RFC models based on multi-class features provided comparatively high efficiency in our study.

Keywords: primary central nervous system lymphoma; predictive modeling; Radiomics; machine learning; PET



Citation: Cui, C.; Yao, X.; Xu, L.; Chao, Y.; Hu, Y.; Zhao, S.; Hu, Y.; Zhang, J. Improving the Classification of PCNSL and Brain Metastases by Developing a Machine Learning Model Based on ¹⁸F-FDG PET. *J. Pers. Med.* **2023**, *13*, 539. <https://doi.org/10.3390/jpm13030539>

Academic Editor: Shengming Deng

Received: 13 January 2023

Revised: 6 March 2023

Accepted: 14 March 2023

Published: 17 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Instruction

The use of imaging techniques to assess brain lesions is crucial in diagnosing and managing neurological disorders. MRI and CT have commonly used imaging modalities,

but they have limited usefulness in providing information on the metabolic activity of brain lesions. In contrast, ^{18}F -FDG PET-CT is a functional imaging modality that can provide valuable information on the metabolic activity of brain lesions, particularly brain tumors.

A review of the literature suggests that 2-deoxy-2-[^{18}F]-fluoro-D-glucose (^{18}F -FDG) PET-CT is a valuable tool in identifying metabolically active brain tumors and monitoring treatment response [1]. Zhao et al. (2014) reported that ^{18}F -FDG PET-CT had high sensitivity and specificity for detecting brain tumors and differentiating them from non-neoplastic lesions [2].

However, the accuracy and usefulness of ^{18}F -FDG PET-CT in CNS diagnosis are still debated among researchers and clinicians. Some studies have reported lower accuracy rates for differentiating between benign and malignant brain tumors. The usefulness of ^{18}F -FDG PET-CT may be affected by the lesion's type and location, surrounding inflammation or edema, and the patient's metabolic state.

Despite these limitations and controversies, the available evidence suggests that ^{18}F -FDG PET-CT remains a valuable tool for assessing brain lesions, particularly in the context of brain tumors. Yang et al. (2019) [3] reported that ^{18}F -FDG PET-CT and MRI had similar diagnostic accuracy in differentiating between high-grade and low-grade gliomas.

Primary central nervous system lymphoma (PCNSL) is a rare type of non-Hodgkin lymphoma that affects the brain, eyes, leptomeninges, or spinal cord. The incidence of PCNSL was 7 cases per 1,000,000 people in the USA in 2013 [4]. The PCNSL accounts for 2–3% of all brain tumors [5] (pp. 971–977). A study reported that the 2-year age-adjusted relative survival rate of PCNSL was 33%, and the corresponding 5-year survival rate of PCNSL was 26% [6]. An accurate diagnosis is crucial for the effective treatment of PCNSL. Currently, combination chemotherapy regimens that include high-dose methotrexate are considered the standard of care for newly diagnosed PCNSL [7]. In contrast, patients with brain metastases require a multidisciplinary approach that involves surgical resection, various radiation treatment modalities, cytotoxic chemotherapy, and targeted molecular treatment [8].

Neuro-imaging using cranial MRI with fluid-attenuated inversion recovery (FLAIR) and T1-weighted sequences before and after contrast injection is the preferred method for diagnosing and monitoring PCNSL [9]. However, distinguishing between PCNSL and brain metastases can be challenging since both present similar MRI signs, such as non-enhancing core and perifocal edema [10]. Moreover, a particular part of PCNSL can demonstrate T2-weighted hyperintensity as other intracranial tumors [11]. ^{18}F -FDG PET can be helpful for differential diagnosis, but it has insufficient specificity [9,12].

Recent, more inspiring studies of MRI-based or PET/CT-based radiomics models combining histograms with texture features have been widely reported in diagnosing and managing glioma and metastases in the brain [13,14]. Nonetheless, due to the low morbidity of PCNSL, the relevant diagnosing model has not been widely investigated yet.

Therefore, we aim to establish several models based on ^{18}F -FDG PET/CT and find an estimator with the best-predicted performance to identify PCNSL to improve diagnosis, affect patients' management, decrease the number of indications to surgical interventions, direct the patient to the most accurate therapy, and, therefore, affect their quality of life.

2. Materials and Methods

Our study follows the guideline, transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) [15]. The statement adhered to the supplement materials as a part of the study (Table S1).

2.1. Study Participants

The study retrospectively reviewed patients with intracranial mass who received an ^{18}F -FDG PET/CT at Jiangsu Cancer Hospital from May 2016 to May 2022. Patients with PCNSL confirmed by pathology and brain metastases confirmed by pathology or clinical history without receiving systemic therapy or brain radiotherapy for the past six months. Due to patients' compliance, the biopsy of brain metastases cannot be feasible for all the

patients whose primary tumor was pathologically confirmed. All the lesions were not postoperative or post-biopsy (Figure 1).

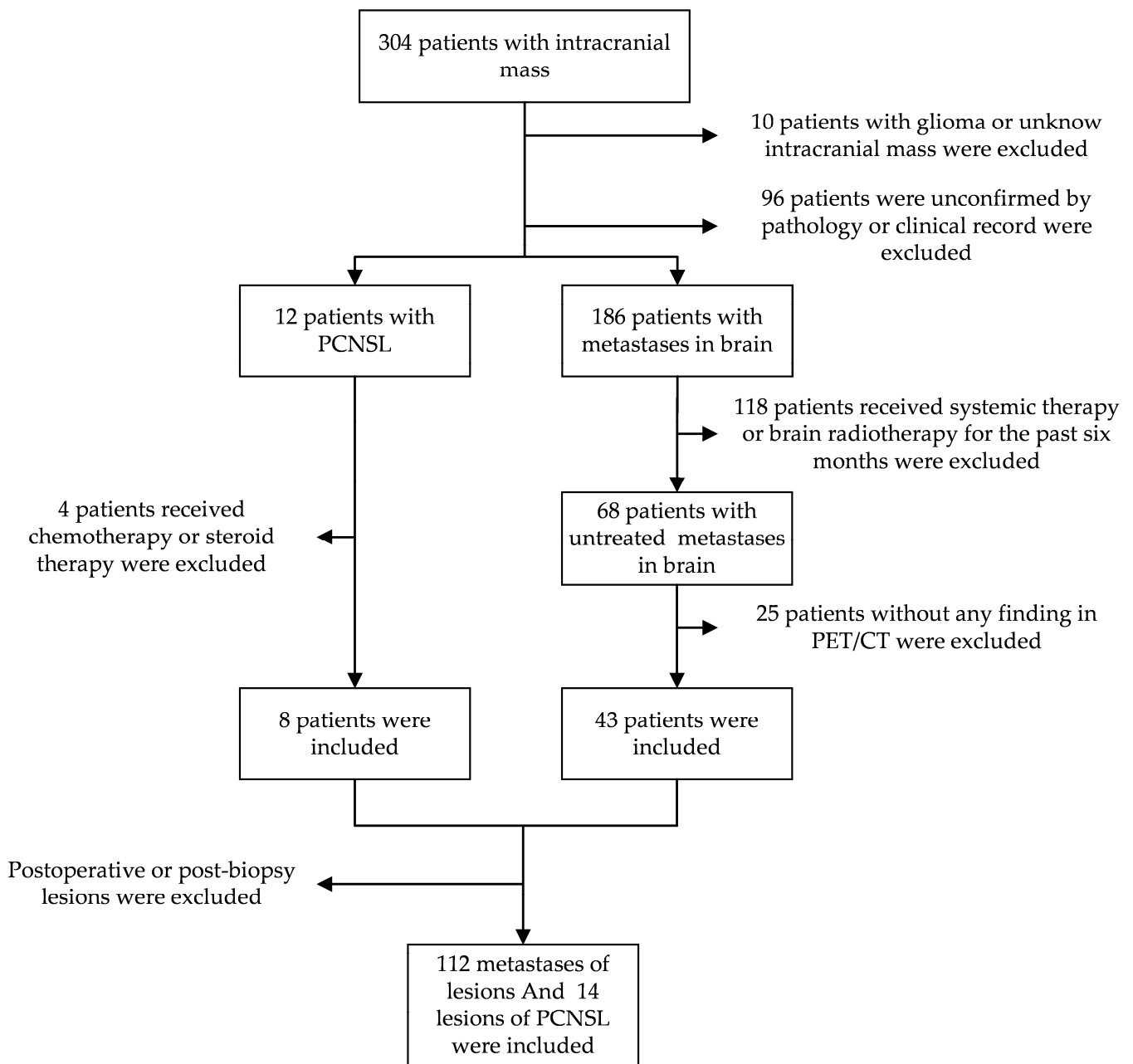


Figure 1. The pathway of inclusion criteria for patients and lesions. One hundred twenty-six lesions were included in the total.

2.2. ¹⁸F-FDG PET/CT Protocol

¹⁸F-FDG PET/CT protocol followed the European Association of Nuclear Medicine’s guidelines [16]. Patients fasted for at least 6 h. The plasma glucose level of all the patients was in a range from 4.0 mmol/L to 8.3 mmol/L. For patients with diabetes, additional restrictions were applied. Only intermediate-acting or short-acting insulin was allowed within 12 h before the administration of ¹⁸F-FDG, and the application of metformin was compromised. The radioactivity of ¹⁸F-FDG for intravenous injection was calculated by body weight, 4.1 ± 0.82 Mbq/kg (range from 2.96 MBq/kg to 5.55 Mbq/kg). The acquisition of the brain starts at 77 ± 2.9 min (range from 74 to 82 min) after ¹⁸F-FDG injection when the PET scan of the torso (from the canthus line to the thigh) was completed.

The brain scan is a separate procedure. The PET/CT (Discover 710 STD GE Healthcare, Waukesha, WI, USA) image acquisition consisted of a 10-min emission scanning with one bed for the brain and low-dose CT for attenuation correction. The voxel size was $3.65 \times 3.65 \times 3.75$ in mm with a matrix of 192×192 . The reconstruction is Vue Point FX with 24 subsets and 2-times iterations. Low-dose CT used 3.75 mm slice thickness, pitch 1.375:1, 140 kV with Auto-mA.

2.3. Segmentation of Images

All the PET/CT images, relevant MRI, and related contrast-enhanced CT images were reviewed using PET VCAR with Integrated Registration, a component of the Advantage Workstation (version 4.6, GE Healthcare, Waukesha, WI, USA).

Segmentation of lesions was performed by two clinical radiologists with over five years of experience. The volume of interest (VOI) was checked by radiology and nuclear medicine physicians with a career in oncological PET/CT interpretation over ten years.

Segmentation of PET volumes was based on the iterative image thresholding method (ITM), which yielded reliable PET volume estimation as previously reported [17]. Relevant MRI and contrast-enhanced CT were used as the reference to adjust the edge of VOIs manually. VOIs were saved and exported as the radiotherapy structure set (RTSS).

2.4. Feature Extraction

All the characters were divided into two groups, the group of density features and the group of multi-class features (Table S2). Briefly, the density-features group contains 10% percentile, 90% percentile, energy, maximum, minimum, and range. The multi-classes-features group includes all first-order characters and the texture characters, such as the gray-level co-occurrence matrix (GLCM), modification of grey-level difference matrix (GLDM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), and neighboring gray-tone difference matrix (NGTDM), 93 features in total.

As the unit of the pixel value is Becquerel per mL, PET images were normalized by the SUV factor Formula (1) and resampled to a uniform voxel size of $2 \times 2 \times 2$ mm³. PyRadiomics (V3.01) (<https://pyradiomics.readthedocs.io/en/latest/index.html>, accessed on 3 May 2022) was used to extract all features [18]. The bin width of 0.5 was derived by dividing the maximum range by 64 [19].

$$SUV\ factor = \frac{W}{D \times 2^{(-t/T)}} \quad (1)$$

Formula (1). W : Body weight (g), D : Injection dose (Bq), t : Delay between injection time and scan time (s), T : Half-life of the isotope (s).

2.5. Model Training and Validation

2.5.1. Statistical Analysis

The present study employed a statistical analysis of three primary steps: resampling, dimensionality reduction, and estimator establishment (Figure 2a). Specifically, to address the issue of imbalanced datasets, the researchers utilized the synthetic minority over-sampling technique and edited nearest neighbors (SMOTEENN) algorithm. SMOTEENN is a hybrid approach that combines the synthetic minority over-sampling technique (SMOTE) and edited nearest neighbors (ENN) algorithms. SMOTE generates synthetic minority class samples to balance the class distribution, while ENN removes examples considered noisy or belonging to the majority class. By combining these two techniques, SMOTEENN can oversample minority class examples and remove potentially noisy or irrelevant examples from the dataset. The tools were provided by imbalanced-learn (Version: 0.9.1) (<https://imbalanced-learn.org/stable/>, accessed on 18 May 2022)

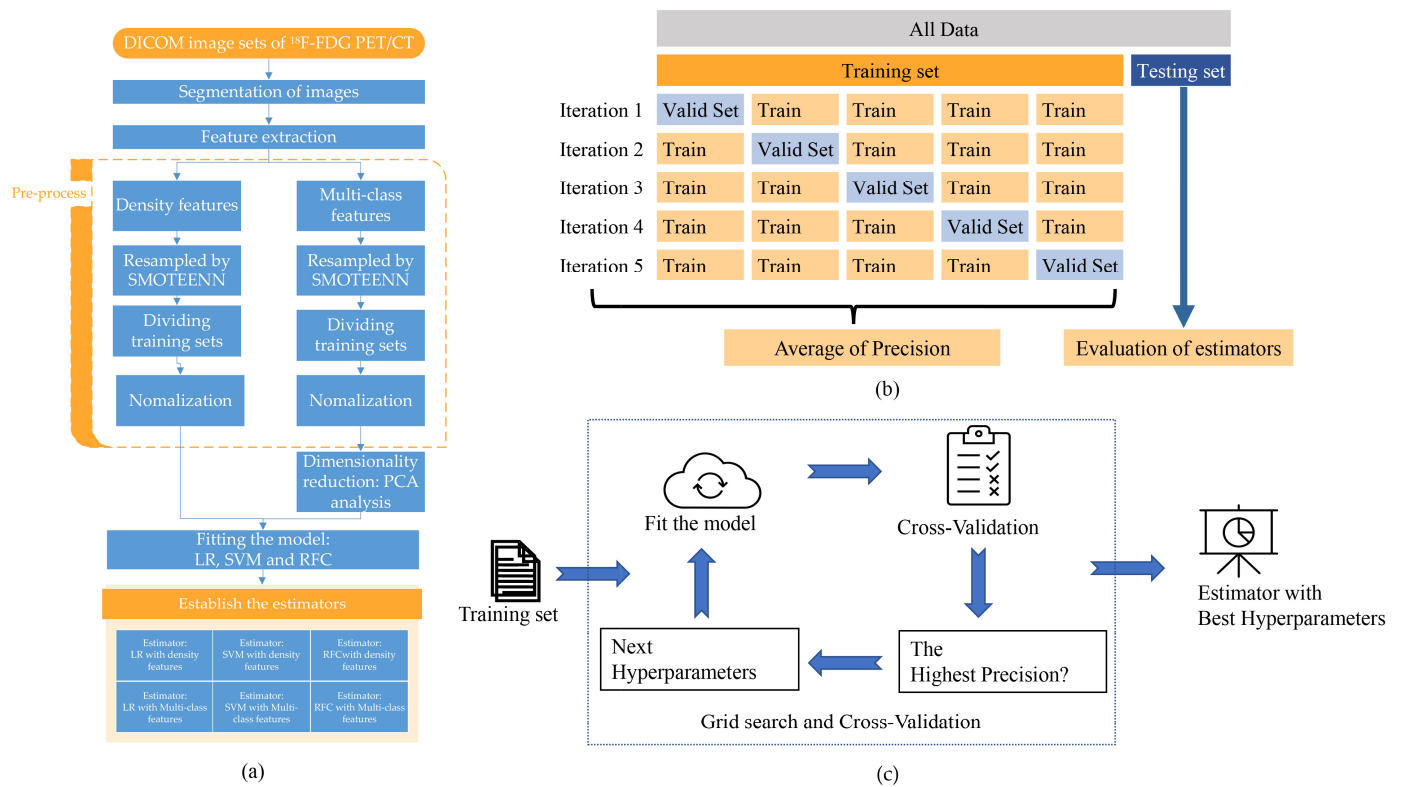


Figure 2. Fit the model and internal cross-validation. (a) The schemes of establishing the estimators. (b) The datasets were split into the training set and the testing set. The training set was divided into five folds for training and cross-validation. (c) The hyperparameters with the highest precision score in cross-validation were chosen.

Principal component analysis (PCA), a linear method known for reducing the dimensions of a dataset while retaining the most relevant information, was employed to achieve the aim mentioned above. The PCA was achieved by transforming the original n-dimensional dataset into a new dataset using an orthogonal transformation [14]. For the last step, three classification algorithms were selected: support vector machine (SVM), logistic regression (LR), and random forest classification (RFC). SVM is a particularly effective classifier for small machine-learning tasks [20]. The LR classifier, while running faster, places greater emphasis on feature engineering [21]. On the other hand, RFC is known to reduce overfitting by averaging decision trees, making it a relatively stable classification method. However, it requires more time to train the model due to its complex calculation process [22]. All the tools above were provided by the scilearn-kit (Version: scikit-learn 1.1.2) (<https://scikit-learn.org/stable/> accessed on 6 August 2022).

2.5.2. Pre-Process of Datasets

Two groups of original datasets were separately resampled by imbalanced-learn (Version: 0.9.1) (<https://imbalanced-learn.org/stable/>, accessed on 18 May 2022). The method of SMOTEENN was used to balance the datasets [23,24].

Two datasets were divided into training sets and testing sets with a ratio of 2:1 using the scilearn-kit (Version: scikit-learn 1.1.2) (<https://scikit-learn.org/stable/>, accessed on 6 August 2022). All of the data were normalized by Standard-Scaler provided by scilearn-kit.

2.5.3. Dimensionality Reduction

Principal component analysis (PCA) was used for dimensionality reduction for the multi-class features group. PCA reduces high-dimensional features into a small number

of principal components (PCs). The PCs will be retained until the cumulative-explained variance is over 0.9.

The dimension of the density-features group was not reduced. Because only six dimensionalities exist in the datasets, dimensionality reduction is unnecessary.

2.5.4. Fitting the Model and Internal Cross-Validation

Two groups of data were fitted to logistic regression (LR), support vector machine (SVM), and random forest classification (RFC) models.

Hyperparameters were determined by grid search with five-fold cross-validation (Figure 2b,c) [25,26]. Briefly, the dataset was split into five folds. In the initial iteration, the first fold was used to validate the model, and the rest folds were used for the training of the model. In the second iteration, the second fold is used as the validation set, while the rest is the training set. This process was repeated five times. The precision Formula (2) of each iteration was averaged. All of the hyperparameters were traversed by grid search. The hyperparameters of each model with the highest precision were selected. Finally, trained by training sets with the best hyperparameters, the six estimators were established from three different models with two data sets.

2.5.5. Evaluation of Estimators

The testing sets and original datasets (the dataset without resampling) were used to evaluate the estimator.

The receiver operating characteristic curve (ROC) with the area under the curve (AUC) is presented. The *F1 score* Formula (2) is a machine-learning metric used in classification models [27]. For imbalanced data, we use the weighted average *F1 score* to compare the efficiency of the estimators.

$$\text{Precision} = \frac{\text{True positives}}{\text{True Positives} + \text{False positives}} \quad (2)$$

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False Negatives}} \quad (3)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Formula (2). The definition of precision (2), recall (3), and Average *F1 score* (4).

3. Result

3.1. Study Participants

The characteristics of patients are demonstrated in Table 1. In total, 8 patients with PCNSL and 43 patients with metastases in the brain were included, with 14 lesions of PCNSL and 112 lesions of metastases in the brain (Figure 1). The primary tumor of all the brain metastases patients was pathologically confirmed. One of the patients, whose primary tumor was adenocarcinoma of the lung, underwent a craniotomy biopsy. Finally, the brain metastases of the lung carcinoma were confirmed. The pathology result of all patients with PCNSL was confirmed by stereotaxic needle biopsy. There is no significant difference in sex and age. The SUVmax of PCNSL and metastases is significantly different.

3.2. Dimensionality Reduction

The study used PCA to project 93 features in the multi-classes-features group to six dimensions. The data of the first three principal components in the training set of the multi-class-features group is shown in Figure 3a. The individual-explained variance ratio and cumulative-explained variance ratio for each principal component are shown in Figure 3b. The cumulative-explained variance ratio of the third principal component is 82.6%, and the sixth is 91.6%, meaning the first 6 principal components contained 91.6% of the information of all 93 features.

Table 1. Characteristics of patients.

Characteristics	PCNSL	Metastases	p Value
Sex			0.0986 ²
Male	3	31	
Female	5	12	
Age	56.00 ± 13.98	59.49 ± 11.74	0.4570 ³
SUVmax ¹	20.14 ± 7.58	12.80 ± 4.84	0.0006 ³
Pathology			
B cell lymphoma	8		
Squamous carcinoma ⁴		12	
Adenocarcinoma ⁴		22	
Melanoma ⁴		3	
Renal clear cell cancer ⁴		2	
Neuroendocrine carcinoma ⁴		2	

¹ The AUC of ROC is 0.78. If the cut-off of SUVmax is 14.42, the sensitivity and specificity are 71.43% and 73.21%, and the F1 score is 0.287. ² The statistical method is Fisher's exact test. ³ The statistical method is the t-test. ⁴ No difference in the average SUVmax among different pathological types was found ($p = 0.5213$) (Figure S1).

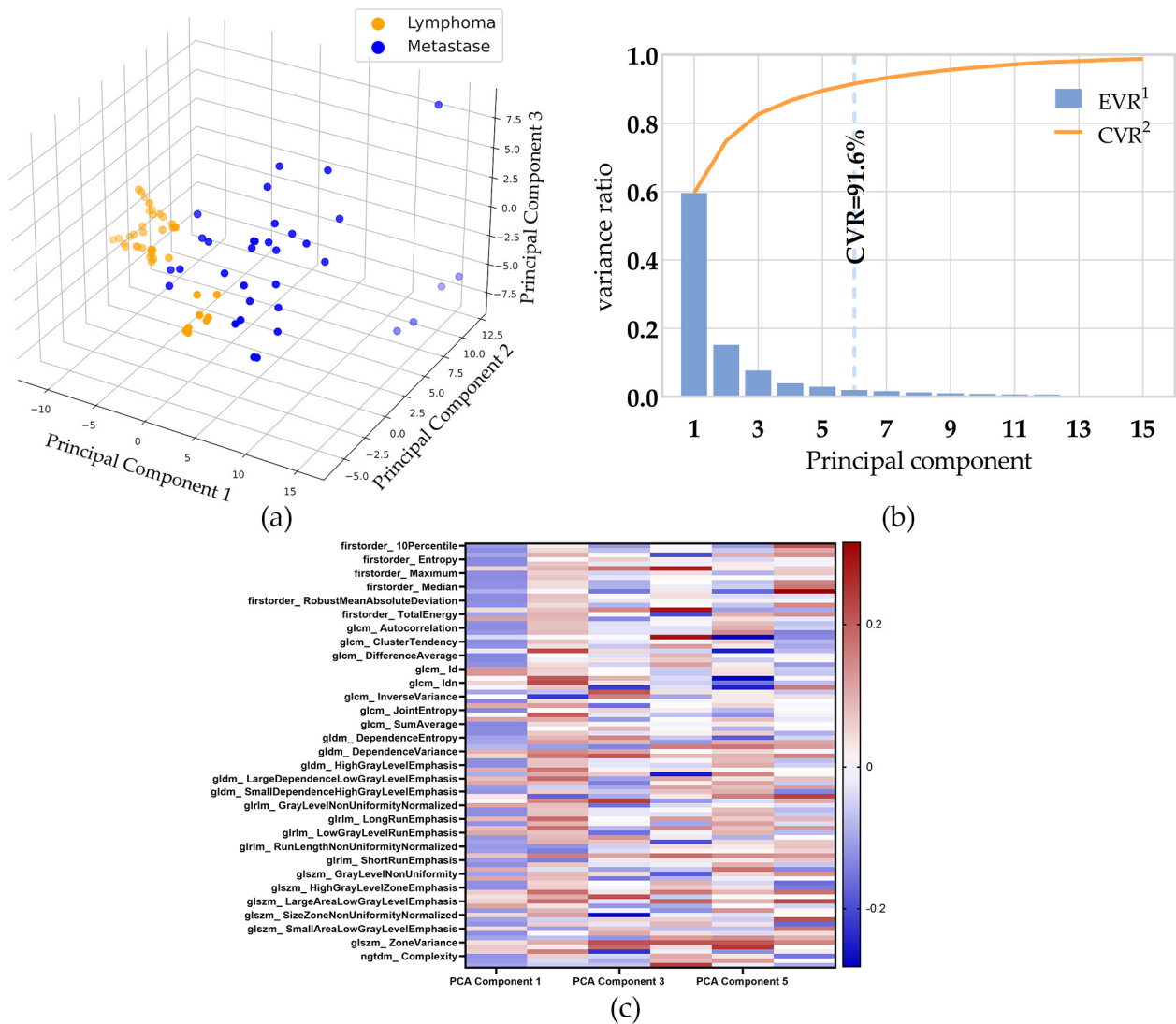


Figure 3. The result PCA. (a) The data of the first three principal components in the multi-class features group training set. (b) The individual-explained variance ratio and cumulative-explained variance ratio for each principal component. (c) The heat map shows a matrix of the PCA loading vectors.¹ EVR: individual-explained variance ratio. ² CVR: cumulative-explained variance ratio.

The PCA loading vectors are shown in Figure 2c and Supplement Table S3. The multi-features dataset was converted from its original dimension to the reduced PCA dimension by using the vectors in the linear transformation.

3.3. Modeling and Validating

3.3.1. Fit the Model and Internal Cross-Validation

The hyperparameters of all the estimators are shown in Table 2.

Table 2. Hyperparameters and precision of estimators.

	Density Features		Multi-Class Features	
	Hyperparameters	Precision	Hyperparameters	Precision
LR	C: 1.0 dual: True multi_class: ovr penalty: l2 solver: liblinear	0.822 ± 0.090	C: 1.4 dual: False multi_class: ovr penalty: l1 solver: liblinear	0.921 ± 0.074
SVM	C: 2.81 gamma: 2.21 kernel: rbf	0.934 ± 0.060	C: 7.01 gamma: 0.21 kernel: poly	1.0 ± 0.0
RFC	bootstrap: False max_depth: 20 max_features: log2 min_samples_leaf: 4 min_samples_split: 16 n_estimators: 500	0.932 ± 0.063	bootstrap: False max_depth: 5 max_features: sqrt min_samples_leaf: 2 min_samples_split: 8 n_estimators: 500.	0.962 ± 0.047

The precision between different models ($p = 0.0137$) and between datasets ($p = 0.0174$) are discrepant. In multiple comparisons between values of precision, only the difference between the SVM model trained by multi-class features and the LR model trained by density features is observed ($p = 0.0025$). The recall of the LR model trained by density features is lower than the others ($p < 0.0001$), while there is no difference was found between the others (Figure 4).

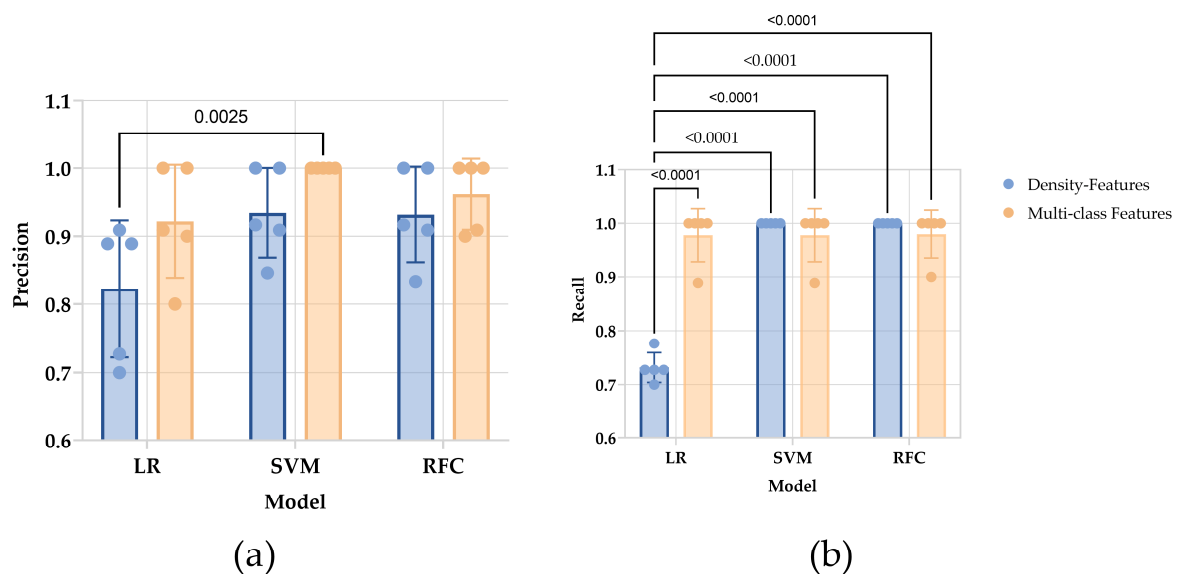


Figure 4. The result of internal cross-validation of chosen estimators. (a) For the precision of each estimator, a significant difference was observed in the SVM model with multi-class features and the LR model with density features. (b) The recall of each estimator. The LR model with density shows the lowest recall. The ANOVA analysis with multiple comparisons calculated p -value.

3.3.2. Evaluation of Estimators

The weighted average *F1 score* of estimators is shown in Table 3. Although all the ROC of estimators shows a nearly perfect performance in the testing set, only SVM and RFC trained by multi-class features exhibit acceptable results, of which the AUCs are 0.92 and 0.93 (>0.9) (Figure 5).

Table 3. The weighted average *F1 scores* of estimators.

	Density Features		Multi-Class Features	
	Testing Set	Original Data	Testing Set	Original Data
LR	0.86	0.79	0.93	0.82
SVM	0.96	0.78	0.98	0.83
RFC	1.00	0.82	0.98	0.85

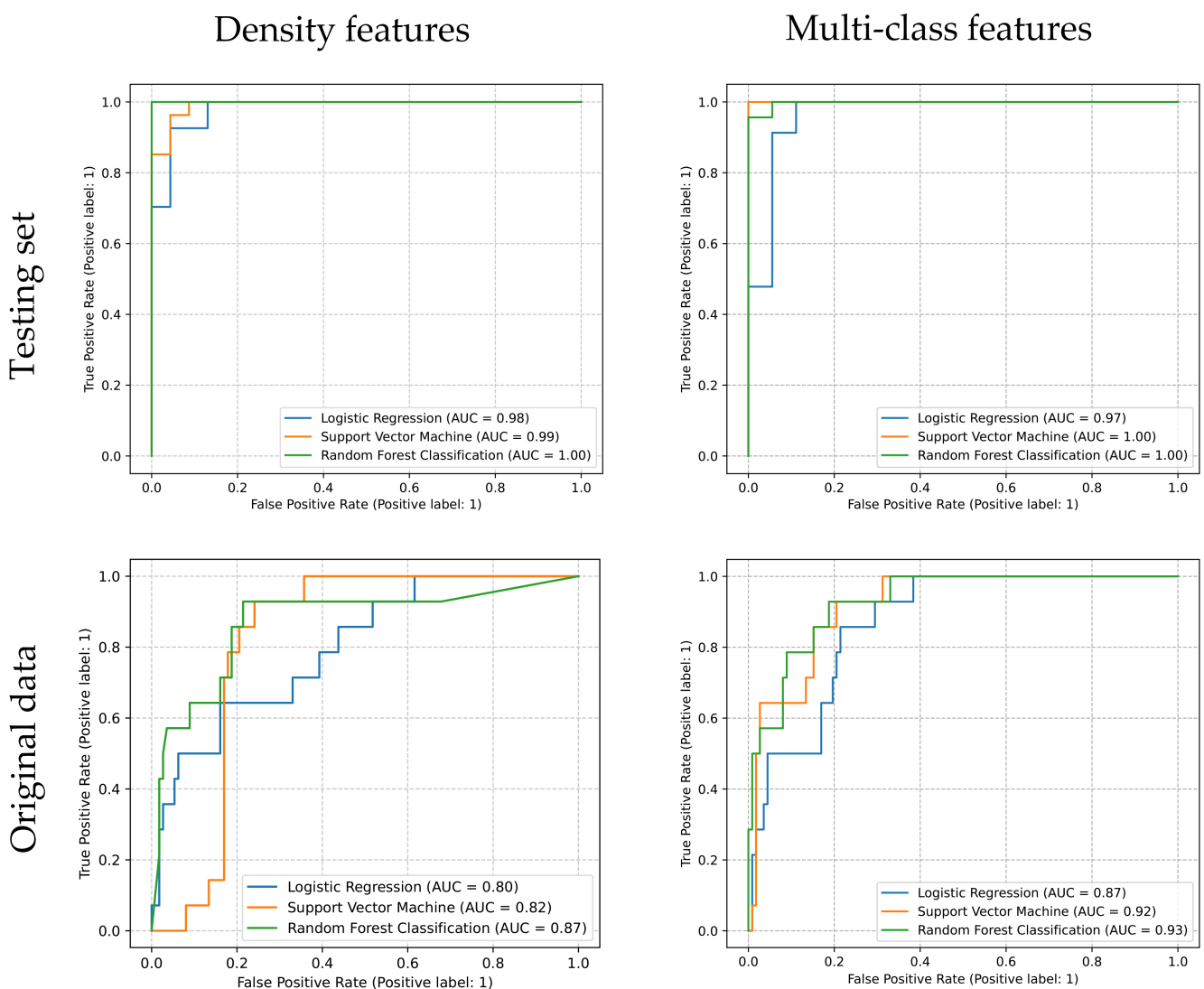


Figure 5. The ROC of all estimators. Tested by original data, the AUC of SVM and RFC trained by multi-class features are 0.92 and 0.93.

4. Discussion

The study established a model to classify PCNSL and neuro-metastases, combining histogram and high-order characteristics from lesions in ^{18}F -FDG PET images. The technique, dimensionality reduction and the balance of data sets, was adopted to reduce the possibility of overfitting.

The SVM and RFC models trained by the multi-class features data set and the RFC models trained by density features show the highest *F1 scores* and AUCs validated by the testing set. However, evaluated by the data sets without resampling, the *F1 scores* and AUCs' reduction of all six estimators can be observed. Nevertheless, the *F1 score* and AUC of the RFC models trained by the multi-class features were still acceptable and relatively higher than others evaluated by the testing and original data set.

^{18}F -FDG PET-CT is a sensitive screening tool for PCNSL patients suspected of systemic involvement [7]. However, A low diagnostic yield of PCNSL for initial staging has been reported [28]. Even if the limitation of ^{18}F -FDG PET in neuro-oncology is widely accepted, some studies argued that the different SUVmax and tumor-normal ratios could be observed in PCNSL and metastases in the brain [12,29]. A similar result can also be drawn from our data; sensitivity and specificity are 71.43% and 73.21%, with a cut-off of 14.42. However, the change SUVmax and tumor-normal ratios may not be conspicuous in atypical PCNSL [30]. Precisely as we noticed, some lesions of PCNSL can be concealed by the high metabolism of the cerebral cortex. In recent years, ^{18}F -FDG PET or MRI-based radiomics features have been reported to distinguish the PCNSL and glioblastoma, which provides a reliable noninvasive method [31–34]. The multi-feature-based diagnosing method should potentially promote the performance in the differential diagnosis between PCNSL and brain metastases. It is just what we discussed in our study to establish a method based on radiomics to increase the diagnosis accuracy of the PCNSL and brain metastases interpreted from ^{18}F -FDG PET.

Due to the disparate incidence of PCNSL and brain metastases, the data set can be highly unbalanced. The incidence of PCNSL was 7 cases per 1,000,000 people in the USA in 2013 [4]. The PCNSL accounts for 2–3% of all brain tumors [5] (pp. 971–977). Relatively, brain metastases develop in approximately 10% to 30% of adults and 6% to 10% of children with cancer [35]. Training with unbalanced datasets may lead to overfitting and underfitting. The synthetic minority over-sampling technique (SMOTE) can be an appropriate option for dealing with imbalanced datasets [24]. The SMOTE is a way to deal with the minority classes in a dataset. This algorithm's fundamental idea is to analyze, simulate, and add the new sample simulated artificially into the original dataset to balance the classes in the original data. In our study, the hyper-sampling method was used. The method combines SOMTE with edited nearest neighbors (ENN), an under-sampling technique that removes the majority class to match the minority class [36]. The method has been used in several clinical studies [37–40].

Actually, for the sure size of the training set, the predictive performance of models decreases with increasing dimensionality [41]. The six visually recognizable features were defined as the group of density. Ninety-three features in the multi-class features group were extracted for PET imaging. The multi-class features can be redundant, and some features can be highly related, which may lead to the over-fitting of the models. It is vital to reduce dimensionality without losing information. PCA determines a set of orthogonal vectors called principal components, defined by a linear combination of the original variables and ordered by the amount of variance explained in component directions [42]. The cumulative-explained variance ratio, the summary of explained variance ratio, has been set to 0.9, which means more than 90 percent of variation from the 93 features has been retained.

In our study, besides the AUCs of ROCs, the weight-average *F1 scores* were used to evaluate the predicted performance of estimators. While ROC was unaffected by skew, precision–recall curves suggest that ROC may mask poor performance [43]. The weight-average *F1 score* is the harmonic mean of precision (also called positive predictive value) and recall (indicated the sensitivity), widely used in information retrieval and information

extraction evaluation [44]. In our study, the weighted average *F1 scores* were used to evaluate the performance of estimators, which calculates the weighted mean of all per-class *F1 scores* while considering each class's support, eliminating the effect of unbalanced data sets.

For the five sixths estimators, the *F1 scores* resulting from the testing set are more prominent than 0.9. The result indicated that precision and sensitivity could be excellent in the testing set. ROC and AUC can also display similar results. In order to evaluate the predicted performance in the real world and the generalization ability of the estimators, we used the original data sets (without resampling) to re-evaluate all estimators. We noticed that all estimators' *F1 scores* or AUCs have a decrease in a certain degree tested by original data sets (without resampling) while considering the class imbalance. Especially in the RFC model trained by density features, the *F1 score* is 1.00 in the testing set and decreases to 0.82 in the original data. We conjecture that overfitting this estimator may decrease the estimators' performance, as reported [45]. On the other hand, the estimator generated from the RFC model trained by multi-class features performs well for both the testing data set and the original data set (without resampling). We conjecture that the characters of the random forests algorithm decrease the possibility of overfitting. Because the random forests deal with the problem of overfitting by creating multiple trees, with each tree trained slightly differently, it overfits differently. The sufficient diagnostic information provided by the multi-class features and the combination of each decision tree offset the effect of overfitting each decision tree.

The low incidence of PCNSL and restricted enrollment criteria restrict the sample size, and further multicenter studies are urgently required. The utilization of various machine learning algorithms has significantly enhanced the efficacy of identifying primary central nervous system lymphoma (PCNSL) and brain metastases. However, it has concurrently augmented the complexity of the practical implementation of these techniques in clinical settings. The random forest model exhibits superior accuracy when dealing with high-dimensional data. Nevertheless, the random forest model's interpretability is greatly diminished by the utilization of multiple decision tree models to determine the final classification outcome through voting.

The ML model based on ^{18}F -FDG can improve the diagnosis of brain lesions by providing clinicians with more precise and consistent information, which can lead to faster and more effective treatment decisions. Radiomics models, which use AI algorithms to analyze medical images, have shown promise in differentiating between brain lesions, including PCNSL and brain metastases.

However, while The ML model based on ^{18}F -FDG has shown potential in improving the diagnosis of brain lesions, more research is needed to fully understand their clinical impact and how to integrate them into clinical practice. Clinicians must be aware of these tools' limitations and potential biases and ensure their use is evidence-based and clinically relevant.

5. Conclusions

The SUVmax of ^{18}F -FDG PET is a proven semi-quantitative indicator; the combination of radiomics and machine learning promotes the performance of PCNLS and brain metastases diagnosis. The *F1 score* and AUC of the RFC model trained by multi-class features are 0.85 and 0.93. The RFC model trained by multi-class features has the potential to revolutionize brain lesions diagnosis and improve patient outcomes. However, they need to integrate into clinical practice cautiously and consider their limitations and biases. More research is needed to fully understand the clinical impact of the model and how it can be best utilized in clinical settings.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jpm13030539/s1>, Figure S1: The SUVmax of brain metastases with different pathology; Table S1: The TRIPOD checklist, Table S2: The features list of the group of density characters and the group of multi-class characters; Table S3: The PCA loading vectors.

Author Contributions: Conceptualization, C.C. and L.X.; methodology, CC; software, C.C. and L.X.; formal analysis, J.Z.; investigation, Y.H. (Yao Hu) and S.Z.; data curation, J.Z.; writing—original draft preparation, C.C. and X.Y.; writing—review and editing, Y.C. and L.X.; visualization, C.C.; supervision, Y.H. (Yuxiao Hu); project administration, C.C. and J.Z.; funding acquisition, C.C., J.Z., Y.H. (Yuxiao Hu); All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Jiangsu Provincial Cancer Hospital Science and Technology Development Fund (No. ZM202018); The Jiangsu Provincial Cancer Hospital Science and Technology Development Fund (No. ZL202214); and The Talents Program of Jiangsu Cancer Hospital (No. YC201801).

Institutional Review Board Statement: The study was conducted under the Declaration of Helsinki and approved by the Ethics Committee of Jiangsu Cancer Hospital (protocol code 2022ke-kuai026 and approved on 4 April 2022).

Informed Consent Statement: Patient consent was waived due to the retrospective nature of the study. The ethics committee has exempted the patient consent. The anonymized ¹⁸F-FDG PET/CT images in research and service development projects.

Data Availability Statement: The data are not publicly available due to institutional data sharing restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pietrzak, A.; Marszałek, A.; Kunikowska, J.; Piotrowski, T.; Medak, A.; Pietrasz, K.; Wojtowicz, J.; Cholewiński, W. Detection of clinically silent brain lesions in [18F]FDG PET/CT study in oncological patients: Analysis of over 10,000 studies. *Sci. Rep.* **2021**, *11*, 18293. [[CrossRef](#)]
2. Zhao, C.; Zhang, Y.; Wang, J. A meta-analysis on the diagnostic performance of (18)F-FDG and (11)C-methionine PET for differentiating brain tumors. *Am. J. Neuroradiol.* **2014**, *35*, 1058–1065. [[CrossRef](#)] [[PubMed](#)]
3. Yang, Y.; He, M.Z.; Li, T.; Yang, X. MRI combined with PET-CT of different tracers to improve the accuracy of glioma diagnosis: A systematic review and meta-analysis. *Neurosurg. Rev.* **2019**, *42*, 185–195. [[CrossRef](#)] [[PubMed](#)]
4. O'Neill, B.P.; Decker, P.A.; Tieu, C.; Cerhan, J.R. The changing incidence of primary central nervous system lymphoma is driven primarily by the changing incidence in young and middle-aged men and differs from time trends in systemic diffuse large B-cell non-Hodgkin's lymphoma. *Am. J. Hematol.* **2013**, *88*, 997–1000. [[CrossRef](#)] [[PubMed](#)]
5. Brastianos, P.K.; Batchelor, T.T. Primary Central Nervous System Lymphoma. In *Encyclopedia of the Neurological Sciences*, 2nd ed.; Aminoff, M.J., Daroff, R.B., Eds.; Academic Press: Oxford, UK, 2014; pp. 971–977. ISBN 978-0-12-385158-1.
6. Puhakka, I.; Kuitunen, H.; Jäkälä, P.; Sonkajarvi, E.; Turpeenniemi-Hujanen, T.; Rönkä, A.; Selander, T.; Korhonen, M.; Kuittinen, O. Primary central nervous system lymphoma high incidence and poor survival in Finnish population-based analysis. *BMC Cancer* **2022**, *22*, 236. [[CrossRef](#)] [[PubMed](#)]
7. Fox, C.P.; Phillips, E.H.; Smith, J.; Linton, K.; Gallop-Evans, E.; Hemmaway, C.; Auer, D.P.; Fuller, C.; Davies, A.J.; McKay, P.; et al. Guidelines for the diagnosis and management of primary central nervous system diffuse large B-cell lymphoma. *Br. J. Haematol.* **2019**, *184*, 348–363. [[CrossRef](#)]
8. Proescholdt, M.A.; Schodel, P.; Doenitz, C.; Pukrop, T.; Hohne, J.; Schmidt, N.O.; Schebesch, K.M. The Management of Brain Metastases-Systematic Review of Neurosurgical Aspects. *Cancers* **2021**, *13*, 1616. [[CrossRef](#)]
9. Scheichel, F.; Marhold, F.; Pinggera, D.; Kiesel, B.; Rossmann, T.; Popadic, B.; Woehrer, A.; Weber, M.; Kitzwoegerer, M.; Geissler, K.; et al. Influence of preoperative corticosteroid treatment on rate of diagnostic surgeries in primary central nervous system lymphoma: A multicenter retrospective study. *BMC Cancer* **2021**, *21*, 754. [[CrossRef](#)]
10. Kuker, W.; Nagele, T.; Korfel, A.; Heckl, S.; Thiel, E.; Bamberg, M.; Weller, M.; Herrlinger, U. Primary central nervous system lymphomas (PCNSL): MRI features at presentation in 100 patients. *J. Neuro-Oncol.* **2005**, *72*, 169–177. [[CrossRef](#)]
11. Haldorsen, I.S.; Espeland, A.; Larsson, E.M. Central nervous system lymphoma: Characteristic findings on traditional and advanced imaging. *Am. J. Neuroradiol.* **2011**, *32*, 984–992. [[CrossRef](#)]
12. Kawai, N.; Miyake, K.; Yamamoto, Y.; Nishiyama, Y.; Tamiya, T. 18F-FDG PET in the diagnosis and treatment of primary central nervous system lymphoma. *Biomed Res. Int.* **2013**, *2013*, 247152. [[CrossRef](#)] [[PubMed](#)]
13. Cao, X.; Tan, D.; Liu, Z.; Liao, M.; Kan, Y.; Yao, R.; Zhang, L.; Nie, L.; Liao, R.; Chen, S.; et al. Differentiating solitary brain metastases from glioblastoma by radiomics features derived from MRI and 18F-FDG-PET and the combined application of multiple models. *Sci. Rep.* **2022**, *12*, 5722. [[CrossRef](#)] [[PubMed](#)]
14. Wang, K.; Qiao, Z.; Zhao, X.; Li, X.; Wang, X.; Wu, T.; Chen, Z.; Fan, D.; Chen, Q.; Ai, L. Individualized discrimination of tumor recurrence from radiation necrosis in glioma patients using an integrated radiomics-based model. *Eur. J. Nucl. Med. Mol.* **2020**, *47*, 1400–1411. [[CrossRef](#)] [[PubMed](#)]

15. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Bmj-Br. Med. J.* **2015**, *350*, g7594. [[CrossRef](#)]
16. Boellaard, R.; Delgado-Bolton, R.; Oyen, W.J.; Giammarile, F.; Tatsch, K.; Eschner, W.; Verzijlbergen, F.J.; Barrington, S.F.; Pike, L.C.; Weber, W.A.; et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: Version 2.0. *Eur. J. Nucl. Med. Mol.* **2015**, *42*, 328–354. [[CrossRef](#)]
17. Jentzen, W.; Freudenberg, L.; Eising, E.G.; Heinze, M.; Brandau, W.; Bockisch, A. Segmentation of PET Volumes by Iterative Image Thresholding. *Soc. Nucl. Med.* **2007**, *48*, 108–114.
18. van Griethuysen, J.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.; Fillion-Robin, J.C.; Pieper, S.; Aerts, H. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [[CrossRef](#)]
19. Orlhac, F.; Soussan, M.; Chouahnia, K.; Martinod, E.; Buvat, I. 18F-FDG PET-Derived Textural Indices Reflect Tissue-Specific Uptake Pattern in Non-Small Cell Lung Cancer. *PLoS ONE* **2015**, *10*, e145063. [[CrossRef](#)]
20. Baesens, B.; Viaene, S.; Van Gestel, T.; Suykens, J.A.K.; Dedene, G.; De Moor, B.; Vanthienen, J. Least Squares Support Vector Machine Classifiers: An Empirical Evaluation. *TEW Res. Rep. 0003* **2000**, 1–16.
21. Menard, S. Six Approaches to Calculating Standardized Logistic Regression Coefficients. *Am. Stat.* **2004**, *58*, 218–223. [[CrossRef](#)]
22. Liaw, A.; Wiener, M.C. Classification and Regression by Randomforest. *R News* **2002**, *2*, 18–22.
23. Tre, G.L.I.; Nogueira, F.; Aridas, C.K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.
24. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
25. Rao, R.B.; Fung, G.; Rosales, R. On the Dangers of Cross-Validation. An Experimental Evaluation. In Proceedings of the SIAM International Conference on Data Mining, Atlanta, GA, USA, 24–26 April 2008; p. 588.
26. Wu, C.; Xue, X.; Song, Y. Research on Cancer Diagnosis Method Based on LightGBM-Gridsearchcv. In Proceedings of the 4th International Conference on Big Data Engineering, Beijing, China, 26–28 May 2022; pp. 122–126.
27. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 29. [[CrossRef](#)]
28. Suh, C.H.; Kim, H.S.; Park, J.E.; Jung, S.C.; Choi, C.G.; Kim, S.J. Primary Central Nervous System Lymphoma: Diagnostic Yield of Whole-Body CT and FDG PET/CT for Initial Systemic Imaging. *Radiology* **2019**, *292*, 440–446. [[CrossRef](#)]
29. Yamaguchi, S.; Hirata, K.; Kobayashi, H.; Shiga, T.; Manabe, O.; Kobayashi, K.; Motegi, H.; Terasaka, S.; Houkin, K. The diagnostic role of (18)F-FDG PET for primary central nervous system lymphoma. *Ann. Nucl. Med.* **2014**, *28*, 603–609. [[CrossRef](#)] [[PubMed](#)]
30. Kawai, N.; Miyake, K.; Okada, M.; Yamamoto, Y.; Nishiyama, Y.; Tamiya, T. Usefulness and limitation of FDG-PET in the diagnosis of primary central nervous system lymphoma. *No Shinkei Geka* **2013**, *41*, 117–126.
31. Kong, Z.; Jiang, C.; Zhu, R.; Feng, S.; Wang, Y.; Li, J.; Chen, W.; Liu, P.; Zhao, D.; Ma, W.; et al. 18F-FDG-PET-based radiomics features to distinguish primary central nervous system lymphoma from glioblastoma. *NeuroImage Clin.* **2019**, *23*, 101912. [[CrossRef](#)] [[PubMed](#)]
32. Kunimatsu, A.; Kunimatsu, N.; Kamiya, K.; Watadani, T.; Mori, H.; Abe, O. Comparison between Glioblastoma and Primary Central Nervous System Lymphoma Using MR Image-based Texture Analysis. *Magn. Reason. Med. Sci.* **2018**, *17*, 50–57. [[CrossRef](#)] [[PubMed](#)]
33. Suh, H.B.; Choi, Y.S.; Bae, S.; Ahn, S.S.; Chang, J.H.; Kang, S.G.; Kim, E.H.; Kim, S.H.; Lee, S.K. Primary central nervous system lymphoma and atypical glioblastoma: Differentiation using radiomics approach. *Eur. Radiol.* **2018**, *28*, 3832–3839. [[CrossRef](#)]
34. Kang, D.; Park, J.E.; Kim, Y.H.; Kim, J.H.; Oh, J.Y.; Kim, J.; Kim, Y.; Kim, S.T.; Kim, H.S. Diffusion radiomics as a diagnostic model for atypical manifestation of primary central nervous system lymphoma: Development and multicenter external validation. *Neuro-Oncology* **2018**, *20*, 1251–1261. [[CrossRef](#)] [[PubMed](#)]
35. Wen, P.Y.; Loeffler, J.S. Management of brain metastases. *Oncology* **1999**, *13*, 941–954. [[PubMed](#)]
36. Batista, G.E.A.P.; Prati, R.C.; Monard, M.C. A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. *SIGKDD Explor. Newsl.* **2004**, *6*, 20–29. [[CrossRef](#)]
37. Wang, H.; Li, X.; Yuan, Y.; Tong, Y.; Zhu, S.; Huang, R.; Shen, K.; Guo, Y.; Wang, Y.; Chen, X. Association of machine learning ultrasound radiomics and disease outcome in triple negative breast cancer. *Am. J. Cancer Res.* **2022**, *12*, 152–164. [[PubMed](#)]
38. Ji, W.; Zhang, Y.; Cheng, Y.; Wang, Y.; Zhou, Y. Development and validation of prediction models for hypertension risks: A cross-sectional study based on 4,287,407 participants. *Front. Cardiovasc. Med.* **2022**, *9*, 928948. [[CrossRef](#)] [[PubMed](#)]
39. Hashimoto-Roth, E.; Surendra, A.; Lavallée-Adam, M.; Bennett, S.; Čuperlović-Culf, M. METAbolomics data Balancing with Over-sampling Algorithms (Meta-BOA): An online resource for addressing class imbalance. *Bioinformatics* **2022**, *38*, 5326–5327. [[CrossRef](#)]
40. Ullah, Z.; Saleem, F.; Jamjoom, M.; Fakieh, B.; Kateb, F.; Ali, A.M.; Shah, B. Detecting High-Risk Factors and Early Diagnosis of Diabetes Using Machine Learning Methods. *Comput. Intell. Neurosc.* **2022**, *2022*, 2557795. [[CrossRef](#)]
41. Oommen, T.; Misra, D.; Twarakavi, N.K.C.; Prakash, A.; Sahoo, B.; Bhandopadhyay, S. An Objective Analysis of Support Vector Machine Based Classification for Remote Sensing. *Math. Geosci.* **2008**, *40*, 409–424. [[CrossRef](#)]
42. Ballabio, D. A MATLAB toolbox for Principal Component Analysis and unsupervised exploration of data structure. *Chemometr. Intell. Lab.* **2015**, *149*, 1–9. [[CrossRef](#)]

43. Jeni, L.A.; Cohn, J.F.; De La Torre, F. Facing Imbalanced Data—Recommendations for the Use of Performance Metrics. In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 2–5 September 2013; pp. 245–251.
44. Takahashi, K.; Yamamoto, K.; Kuchiba, A.; Koyama, T. Confidence interval for micro-averaged F 1 and macro-averaged F 1 scores. *Appl. Intell.* **2022**, *52*, 4961–4972. [[CrossRef](#)]
45. Kernbach, J.M.; Staartjes, V.E. Foundations of Machine Learning-Based Clinical Prediction Modeling: Part II—Generalization and Overfitting. *Acta Neurochir. Suppl.* **2022**, *134*, 15–21. [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.