Bridging Structural and Functional Imaging: Integrated PET/CT Radiomics with Explainable Machine Learning

Pierfrancesco Novielli[1,3,*], Donato Romano[1,3], Michele Magarelli[1], Pierpaolo Di Bitonto[1,3], Roberto Bellotti[2,3] and Sabina Tangaro[1,3,*]

[1]Dipartimento di Scienze del Suolo, della Pianta e degli Alimenti, Università degli Studi di Bari Aldo Moro, Bari, Italy
[2]Dipartimento Interateneo di Fisica "M. Merlin", Università degli Studi di Bari Aldo Moro, Bari, Italy
[3]Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Bari, Italy

Abstract

This work presents a radiomic-based framework for the classification of metastatic versus non-metastatic lung lesions using multimodal imaging features from PET and CT scans. The proposed pipeline integrates Extreme Gradient Boosting (XGBoost) with SHapley Additive exPlanations (SHAP) to enhance model interpretability. A comparative evaluation shows that the fusion of PET and CT radiomics improves classification performance compared to unimodal analysis. SHAP values provide insights into feature importance, highlighting the complementary roles of structural and metabolic imaging. This Late-Breaking Work addresses key challenges in explainability, reproducibility, and clinical relevance, contributing toward trustworthy AI in medical imaging.

Keywords

Explainable AI, SHAP, Radiomics, PET-CT Imaging, Lung Cancer Classification, Machine Learning

1. Introduction

Lung cancer and metastatic pulmonary lesions represent a persistent challenge in clinical oncology. Accurate and early diagnosis is essential to guide treatment strategies and improve patient survival [1]. Imaging modalities such as Computed Tomography (CT) and Positron Emission Tomography (PET) have long been employed for this purpose, offering insights into the structural and functional properties of lung lesions. However, conventional imaging assessment often relies on subjective visual interpretation, which may lead to diagnostic variability and uncertainty.

Radiomics has emerged as a promising approach to overcome these limitations by extracting quantitative descriptors from medical images [2]. These features—spanning shape, intensity, and texture—can capture subtle imaging patterns that may not be visible to the naked eye. When combined with machine learning (ML), radiomics enables data-driven classification models that support decision-making and improve diagnostic reproducibility.

A clinically relevant application of radiomics is in the characterization of solitary pulmonary nodules (SPNs), small rounded lesions (≤3 cm) frequently identified in routine imaging [3]. Differentiating benign from malignant SPNs remains a complex task. Benign nodules often exhibit smooth contours or calcifications, whereas malignant nodules may present irregular borders and faster growth rates [4, 5]. Although histological analysis provides definitive diagnosis, it is invasive and not always feasible. Therefore, non-invasive biomarkers derived from PET and CT images play a critical role in patient stratification [6].

CT offers high-resolution morphological details such as lesion shape, density, and anatomical context [7]. PET, using fluorodeoxyglucose (FDG), measures glucose metabolism and highlights areas of increased uptake typically associated with malignancy [8]. While both modalities offer valuable but

Late-breaking work, Demos and Doctoral Consortium, colocated with the 3rd World Conference on eXplainable Artificial Intelligence: July 09–11, 2025, Istanbul, Turkey
*Corresponding author.
pierfrancesco.novielli@uniba.it (P. Novielli); sabina.tangaro@uniba.it (S. Tangaro)
0000-0001-8773-0636 (P. Novielli); 0009-0002-3828-0469 (D. Romano); 0009-0008-9119-5638 (M. Magarelli); 0000-0001-9940-4002 (P. D. Bitonto); 0000-0003-3198-2708 (R. Bellotti); 0000-0002-1372-3916 (S. Tangaro)

distinct information, PET may produce false positives due to inflammatory uptake, and CT findings can be non-specific. In this context, the integration of radiomic features extracted from both PET and CT scans allows for a more comprehensive characterization of lesion properties. Such structural-functional fusion has shown promise in improving diagnostic accuracy, particularly in oncologic applications [9].

In this work, we present a machine learning-based framework that combines radiomic features from PET, CT, and PET-CT images to classify lung lesions as metastatic or non-metastatic. We use eXtreme Gradient Boosting (XGBoost) for its efficiency and strong performance on structured data [10]. Radiomic features extracted from each imaging modality include intensity metrics (e.g., SUVmin, SUVmean), shape descriptors, and texture-based features such as those derived from the Gray-Level Co-occurrence Matrix (GLCM).

A key objective of this study is to address the issue of model interpretability. While ML models can achieve high accuracy, their clinical adoption is limited if predictions are not explainable. To this end, we integrate Explainable Artificial Intelligence (XAI) into our workflow, specifically through SHapley Additive exPlanations (SHAP) [11]. SHAP assigns importance values to each feature based on its contribution to the model output, offering both global and patient-specific interpretability [12, 13].

The main contributions of this study are:

- A comparative analysis of radiomic features derived from CT, PET, and their integration through feature-level fusion.
- The use of SHAP to identify and visualize the most influential features contributing to lesion classification.
- A transparent and reproducible pipeline that combines predictive accuracy with clinical interpretability.

By bridging radiomics, machine learning, and explainability, this work contributes to the development of reliable AI tools for lung cancer diagnosis. The proposed framework supports not only performance benchmarking across modalities but also delivers insights into feature relevance, reinforcing clinical trust and interpretability in AI-based imaging solutions.

2. Related Work

Radiomics has emerged as a powerful approach for quantifying imaging biomarkers, with early landmark studies such as Aerts et al. [14] demonstrating its utility in predicting prognosis in lung cancer using CT-based features. Subsequent works extended radiomic analysis to PET and PET-CT imaging, highlighting the potential of metabolic features for tumor characterization [9, 15].

Several studies have explored the predictive power of radiomics in distinguishing benign from malignant lung lesions [16, 17]. However, most models have prioritized performance over interpretability, limiting their clinical applicability.

Explainable AI techniques, such as SHAP, have recently gained attention in the biomedical domain, offering insights into feature contributions at both global and local levels. Compared to other XAI methods like LIME or Grad-CAM, SHAP is particularly well-suited for tabular radiomic data thanks to its stability, local consistency, and its solid mathematical foundation based on cooperative game theory [11].

Radiomic features are typically structured, high-dimensional, and exhibit complex correlations, making SHAP's local accuracy and ability to model feature interactions particularly valuable for deriving reliable and clinically meaningful interpretations[18].

While some recent works have applied SHAP to CT or PET-based radiomic classifiers [19], few have systematically investigated its application to multimodal PET-CT fusion. To the best of our knowledge, this is among the first studies to combine PET and CT radiomic features with SHAP-based interpretation for metastatic lung lesion classification, offering a transparent and reproducible pipeline that balances performance and explainability.

3. Materials and Methods

3.1. Dataset Description

This study employs a publicly available dataset from Kirienko et al. [9], which includes patients who underwent FDG PET/CT scans for the evaluation of lung lesions between 2011 and 2017. Eligibility criteria required patients to be over 18 years old and to have a histologically confirmed diagnosis of either primary or metastatic lung tumors.

Importantly, the dataset is already provided in a structured format containing radiomic features that were pre-extracted using the LIFEx software package from semiautomatically segmented PET and CT images [20]. Clinical metadata such as age, sex, and histological subtype were also available. Further demographic details are reported in the original publication.

3.2. Data Preparation

For this analysis, we selected the 468 patients who underwent both PET and CT imaging. Among them, 105 were diagnosed with metastatic lesions and 363 with non-metastatic lesions. Radiomic features were extracted from segmented lung lesions using standardized protocols, generating three datasets:

- **CT-only**: 41 radiomic features.
- **PET-only**: 43 radiomic features.
- **PET+CT**: 84 combined features.

3.3. Addressing Class Imbalance

The dataset was affected by a pronounced class imbalance. To mitigate this, we applied random undersampling of the majority class (non-metastatic), generating balanced subsets of 105 vs. 105 samples. This process was repeated 100 times with different random seeds to ensure coverage and reduce sampling bias. We also experimented with SMOTE, but it resulted in lower performance and increased overfitting. Hence, undersampling was adopted for its better generalization performance in cross-validation [21].

3.4. Machine Learning Approach

Our classification pipeline is outlined in Figure 1. The XGBoost algorithm was chosen after a comparative evaluation against Random Forest, Support Vector Machines (SVM), and Logistic Regression. Across all datasets, XGBoost consistently achieved the highest AUROC and F1-score. Furthermore, its compatibility with SHAP makes it particularly suitable for interpretable radiomic applications [10].

Each model was trained using 10-fold cross-validation, repeated 100 times to ensure robustness. Evaluation metrics included F1-score, Accuracy, Specificity, Precision, Sensitivity, Area Under the ROC Curve (AUROC), and Area Under the Precision-Recall Curve (AUPRC).

3.5. Explainable AI with SHAP

To interpret model predictions, we employed SHapley Additive exPlanations (SHAP), a game-theoretic approach that assigns each feature a contribution score for every prediction [11]. SHAP values were computed for all models and used to analyze feature importance both globally and locally.

The SHAP value for a feature j in a sample x is defined as:

$$\Phi_j(x) = \sum_{F \subseteq S - j} \frac{|F|!(|S| - |F| - 1)!}{|S|!} [f_x(F \cup j) - f_x(F)] \tag{1}$$

This enables the identification of radiomic traits most responsible for distinguishing between metastatic and non-metastatic lesions, offering actionable insights to clinicians and radiologists.

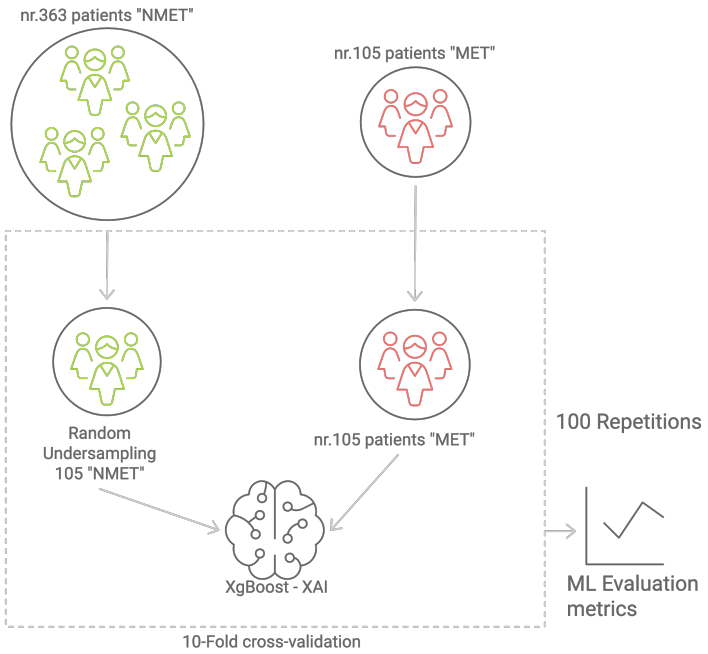


Figure 1: Pipeline of the radiomics classification workflow. Due to class imbalance (105 metastatic vs. 363 non-metastatic cases), the dataset was balanced using random undersampling of the majority class. This process was repeated 100 times to generate multiple balanced subsets, ensuring robustness and reducing sampling bias. For each subset, a 10-fold cross-validation was performed, and final performance metrics were averaged across all repetitions.

4. Results

This study evaluated the effectiveness of radiomic features extracted from CT, PET, and the integration of features from both modalities in classifying metastatic versus non-metastatic lung lesions using the XGBoost classifier. Results demonstrated that PET-derived features provided higher discriminative power compared to CT features alone, and the integration of features from both modalities yielded the best performance.

4.1. Classification Performance

The model was evaluated using a repeated 10-fold cross-validation strategy. Classification metrics, including Accuracy, F1-score, Precision, Sensitivity, Specificity, AUROC, and AUPRC, were averaged over 100 repetitions (Table 1). PET-based features achieved the highest standalone performance (AUROC: 0.863 ± 0.053), while combining PET and CT features further improved accuracy and robustness (AUROC: 0.887 ± 0.048).

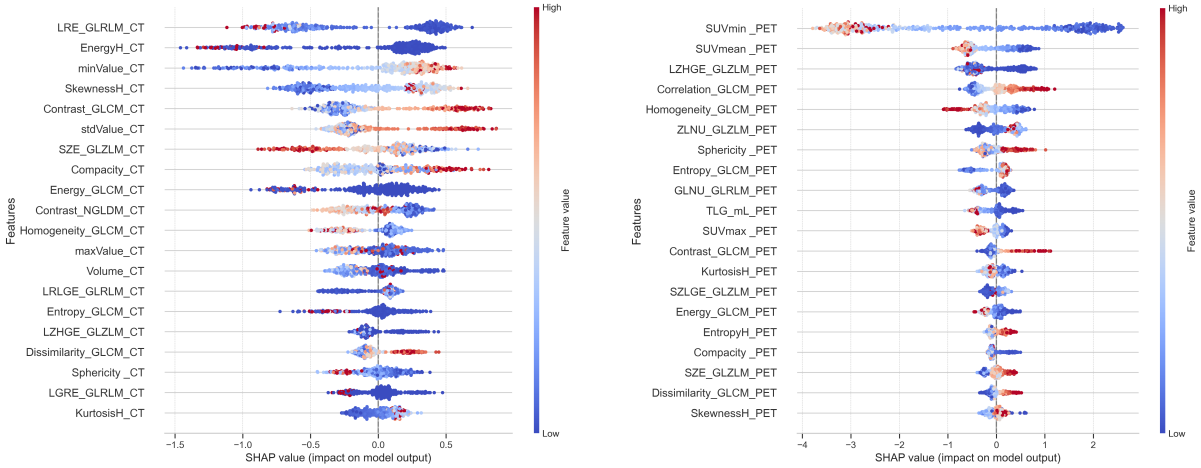
Metric	CT	PET	PET-CT
F1 Score	0.679 ± 0.068	0.802 ± 0.052	0.828 ± 0.054
Accuracy	0.674 ± 0.063	0.797 ± 0.053	0.824 ± 0.054
Specificity	0.655 ± 0.090	0.767 ± 0.085	0.798 ± 0.078
Precision	0.670 ± 0.064	0.785 ± 0.062	0.811 ± 0.062
Sensitivity	0.694 ± 0.098	0.827 ± 0.079	0.850 ± 0.078
AUROC	0.728 ± 0.067	0.863 ± 0.053	0.887 ± 0.048
AUPRC	0.707 ± 0.080	0.825 ± 0.076	0.865 ± 0.067

Table 1
Performance metrics of the XGBoost classifier for each imaging modality, averaged over repeated 10-fold cross-validation.

4.2. Explainability with SHAP

To provide interpretability to the model's decisions, SHAP (SHapley Additive exPlanations) analysis was performed for each imaging modality. SHAP summary plots visualize both the importance of individual features and the direction of their influence on predictions. In each plot, features are ranked by their overall contribution to the model's output. Each dot represents a patient; its color reflects the actual feature value (red = high, blue = low), and its horizontal position indicates the SHAP value—i.e., how much that feature increases (right) or decreases (left) the probability of predicting a metastasis.

Figure 2 presents the SHAP summary plots for the CT, PET, and PET–CT feature integration datasets, highlighting both modality-specific predictors and complementary features contributing to metastasis classification.



(a) CT-based radiomics features

(b) PET-based radiomics features

(c) Structural-functional radiomic feature integration (CT + PET)

Figure 2: SHAP summary plots for the classification of metastatic versus non-metastatic lung lesions using (a) CT-based, (b) PET-based, and (c) integrated CT-PET radiomic features. Each dot represents a patient. The x-axis indicates the SHAP value (feature impact), while the color represents the actual feature value (red = high, blue = low). Features are ordered by their average contribution to the model's predictions. Positive SHAP values indicate a shift toward metastasis classification, while negative values support a non-metastatic prediction.

CT-based radiomics features. In the CT dataset, texture and intensity-based features were the primary contributors. Long Run Emphasis (LRE_GLRLM_CT) and High-Order Energy (EnergyH_CT)

exhibited a negative association with metastasis probability, suggesting that metastatic lesions tend to be more heterogeneous and have a less uniform intensity distribution. Additionally, shape descriptors such as Compacity_CT contributed to capturing geometric variability between lesion types.

PET-based radiomics features. Metabolic and texture-related PET features strongly influenced model predictions. SUVmin_PET emerged as the most influential feature and, interestingly, exhibited a negative correlation with metastasis, suggesting that some metastatic lesions may present lower focal uptake. SUVmean_PET and Correlation_GLCM_PET also emerged as key features, highlighting complex uptake patterns and spatial relationships indicative of malignancy.

Integrated CT-PET features. The integration of PET and CT features enabled complementary information fusion. SUVmin_PET remained the most influential predictor, reinforcing its discriminative value. Features such as LZHGE_GLZLM_PET (Large Zone High Gray-Level Emphasis) captured additional textural complexity in metastatic lesions, while EnergyH_CT contributed morphological information, confirming the benefit of multimodal radiomic fusion in enhancing model explainability and performance.

5. Discussion

This study highlights the potential of radiomic features extracted from PET and CT imaging for the non-invasive classification of lung lesions as metastatic or non-metastatic. PET-derived metabolic features, such as *SUVmin* and *SUVmean*, were among the most discriminative. Interestingly, *SUVmin* was negatively correlated with metastasis, suggesting that certain metastatic lesions may exhibit lower metabolic activity. This finding may reflect underlying biological factors, such as tumor heterogeneity or microenvironmental effects, and supports previous work emphasizing the complex behavior of radiomic biomarkers [22].

Texture-based features, particularly those derived from Gray-Level Co-occurrence Matrix (GLCM) and Gray-Level Zone Length Matrix (GLZLM), also showed strong associations with metastatic status. Metrics such as *Correlation_GLCM_PET*, *Contrast_GLCM_CT*, and *LZHGE_GLZLM_PET* indicated greater heterogeneity in metastatic lesions, consistent with prior findings on texture complexity in malignant tissues [23].

Multimodal integration of PET and CT features significantly improved classification performance over single-modality models. PET provided insights into metabolic activity, while CT added valuable information about lesion morphology and signal distribution. The complementary nature of these modalities reinforces the growing interest in multi-parametric imaging for precision oncology [24].

Importantly, SHAP-based analysis enabled interpretable model outputs, identifying the most influential features and their direction of impact on classification. This enhances transparency, which is a critical factor for clinical adoption of AI tools [25]. By highlighting feature contributions on a per-sample basis, SHAP promotes model trustworthiness and facilitates integration into decision-support systems.

Despite these promising results, several limitations should be acknowledged. The dataset, though carefully curated, is relatively small, which may affect generalizability. Although class imbalance was addressed via repeated random undersampling, alternative approaches such as class-weighted loss functions could be considered in future work to further mitigate bias. Moreover, some features that were predictive in the unimodal CT or PET settings became less relevant when combined, suggesting possible redundancy or dominance effects in the fused feature space—an aspect that deserves deeper exploration, potentially with the aid of domain knowledge. Finally, while SHAP provided interpretable insights, its limitations—particularly in the presence of highly correlated features—should not be overlooked [26]. Alternative or complementary XAI methods, such as counterfactuals or interaction-aware attributions, may help further refine the interpretability and clinical utility of radiomic models.

6. Conclusions

This work presents a preliminary investigation into the use of radiomic features extracted from PET and CT imaging for classifying lung lesions as metastatic or non-metastatic. The integration of metabolic and structural descriptors allowed for a more comprehensive characterization of lesion properties, offering insights into how tumor biology affects both tissue morphology and metabolic behavior. While the combination of features did not yield a drastic performance gain over PET alone, it provided a richer and more interpretable representation of disease traits.

The use of XGBoost, paired with SHAP-based explanations, enabled transparent feature attribution and highlighted complementary roles of CT and PET-derived metrics. These insights could support a better understanding of disease patterns and assist clinicians in identifying key imaging phenotypes associated with metastatic progression.

Although the findings are promising, this study represents an early step toward clinically applicable radiomic decision support. Future work will explore deep learning models capable of learning complex feature hierarchies directly from imaging data, as well as more advanced explainability tools. In particular, the use of SHAP interaction values may reveal synergistic relationships between features, shedding light on how structural and metabolic traits jointly influence classification.

Validation on external, multi-institutional cohorts will be critical to assess generalizability. Ultimately, interpretable and biologically informed AI frameworks have the potential to enhance diagnostic accuracy and support precision oncology workflows in lung cancer care.

Acknowledgments

Authors would like to thank the resources made available by ReCaS, a project funded by the MIUR (Italian Ministry for Education, University and Research) in the "PON Ricerca e Competitività 2007–2013-Azione I-Interventi di rafforzamento strutturale" PONa3_00052, Avviso 254/Ric, University of Bari.

Declaration on Generative AI

During the preparation of this work, the authors used OpenAI's GPT-4 (ChatGPT) exclusively for grammar and spelling checks. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] S. J. Adams, E. Stone, D. R. Baldwin, R. Vliegenthart, P. Lee, F. J. Fintelmann, Lung cancer screening, The Lancet 401 (2023) 390–408.

[2] A. A. Ardakani, N. J. Bureau, E. J. Ciaccio, U. R. Acharya, Interpretation of radiomics features–a pictorial review, Computer methods and programs in biomedicine 215 (2022) 106609.

[3] H. T. Winer-Muram, The solitary pulmonary nodule, Radiology 239 (2006) 34–49.

[4] A. Câmara-de Souza, M. Toyoshima, M. Giannella, D. Freire, C. Camacho, D. Lourenço Jr, M. Rocha, T. Bacchella, R. Jureidini, M. Machado, et al., Insulinoma: A retrospective study analyzing the differences between benign and malignant tumors, Pancreatology 18 (2018) 298–303.

[5] S. H. Bradley, B. S. Bhartia, M. E. Callister, W. T. Hamilton, N. L. F. Hatton, M. P. Kennedy, L. T. Mounce, B. Shinkins, P. Wheatstone, R. D. Neal, Chest x-ray sensitivity and lung cancer outcomes: a retrospective observational study, British Journal of General Practice 71 (2021) e862–e868.

[6] K. M. Latimer, T. F. Mott, Lung cancer: diagnosis, treatment principles, and screening, American family physician 91 (2015) 250–256.

[7] T. M. Buzug, Computed tomography, in: Springer handbook of medical technology, Springer, 2011, pp. 311–342.

[8] D. L. Bailey, M. N. Maisey, D. W. Townsend, P. E. Valk, Positron emission tomography, volume 2, Springer, 2005.
[9] M. Kirienko, L. Cozzi, A. Rossi, E. Voulaz, L. Antunovic, A. Fogliata, A. Chiti, M. Sollini, Ability of fdg pet and ct radiomics features to differentiate between primary and metastatic lung lesions, European journal of nuclear medicine and molecular imaging 45 (2018) 1649–1660.
[10] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.
[11] S. Lundberg, A unified approach to interpreting model predictions, arXiv preprint arXiv:1705.07874 (2017).
[12] P. Novielli, D. Romano, M. Magarelli, D. Diacono, A. Monaco, N. Amoroso, M. Vacca, M. De Angelis, R. Bellotti, S. Tangaro, Personalized identification of autism-related bacteria in the gut microbiome using explainable artificial intelligence, Iscience 27 (2024).
[13] P. Novielli, M. Magarelli, D. Romano, L. de Trizio, P. Di Bitonto, A. Monaco, N. Amoroso, A. M. Stellacci, C. Zoani, R. Bellotti, et al., Climate change and soil health: Explainable artificial intelligence reveals microbiome response to warming, Machine Learning and Knowledge Extraction 6 (2024) 1564–1578.
[14] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nature communications 5 (2014) 4006.
[15] S. Ha, H. Choi, J. C. Paeng, G. J. Cheon, Radiomics in oncological pet/ct: a methodological overview, Nuclear medicine and molecular imaging 53 (2019) 14–29.
[16] Y. Zheng, D. Zhou, H. Liu, M. Wen, Ct-based radiomics analysis of different machine learning models for differentiating benign and malignant parotid tumors, European radiology 32 (2022) 6953–6964.
[17] M. P. Starmans, R. L. Miclea, S. R. Van Der Voort, W. J. Niessen, M. G. Thomeer, S. Klein, Classification of malignant and benign liver tumors using a radiomics approach, in: Medical Imaging 2018: Image Processing, volume 10574, SPIE, 2018, pp. 343–349.
[18] Y. Wang, J. Lang, J. Z. Zuo, Y. Dong, Z. Hu, X. Xu, Y. Zhang, Q. Wang, L. Yang, S. T. Wong, et al., The radiomic-clinical model using the shap method for assessing the treatment response of whole-brain radiotherapy: a multicentric study, European Radiology 32 (2022) 8737–8747.
[19] F. Duan, M. Zhang, C. Yang, X. Wang, D. Wang, Non-invasive prediction of lymph node metastasis in nsclc using clinical, radiomics, and deep learning features from 18f-fdg pet/ct based on interpretable machine learning, Academic Radiology 32 (2025) 1645–1655.
[20] C. Nioche, F. Orlhac, S. Boughdad, S. Reuzé, J. Goya-Outi, C. Robert, C. Pellot-Barakat, M. Soussan, F. Frouin, I. Buvat, Lifex: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity, Cancer research 78 (2018) 4786–4789.
[21] N. Qazi, K. Raza, Effect of feature selection, smote and under sampling on class imbalance classification, in: 2012 UKSim 14th international conference on computer modelling and simulation, IEEE, 2012, pp. 145–150.
[22] L. Pan, P. Gu, G. Huang, H. Xue, S. Wu, Prognostic significance of suv on pet/ct in patients with esophageal cancer: a systematic review and meta-analysis, European journal of gastroenterology & hepatology 21 (2009) 1008–1015.
[23] Z. Wang, A. Zheng, Y. Li, W. Dong, X. Liu, W. Yuan, F. Gao, X. Duan, 18f-psma-1007 pet/ct performance on risk stratification discrimination and distant metastases prediction in newly diagnosed prostate cancer, Frontiers in Oncology 11 (2021) 759053.
[24] S. R. Cherry, Multimodality imaging: Beyond pet/ct and spect/ct, in: Seminars in nuclear medicine, volume 39, Elsevier, 2009, pp. 348–353.
[25] A. Ghosh, D. Kandasamy, Interpretable artificial intelligence: why and when, American Journal of Roentgenology 214 (2020) 1137–1138.
[26] D. Fryer, I. Strümke, H. Nguyen, Shapley values for feature selection: The good, the bad, and the axioms, Ieee Access 9 (2021) 144352–144360.