Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification of bone marrow tumor burden in relapsed/refractory acute lymphoblastic leukemia using 18F-FDG PET/CT radiomics: a multicenter study

bingyan zhang

zhangbingyan108@163.com

Capital Medical University Affiliated Beijing Friendship Hospital

LiJing Wei

Beijing Boren Hospital, Beijing, China

Xiao Lei

Beijing Boren Hospital, Beijing, China

Shuang Yao

Beijing Fengtai You'anmen Hospital, Beijing, China.

Chao Wang

SinoUnion Healthcare Inc., Beijing, China

Hui Zhang

Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China

JiGang Yang

Capital Medical University Affiliated Beijing Friendship Hospital https://orcid.org/0000-0002-0538-7676

Research Article

Keywords: Acute lymphoblastic leukemia, FDG, PET/CT, Radiomics, Minimal residual disease

Posted Date: June 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9956377/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Purpose

To develop a machine learning-based model for differentiating between minimal residual disease (MRD)-positive and MRD-negative patients with acute lymphoblastic leukemia (ALL), and for stratifying MRD-positive patients into low-burden and high-burden subgroups.

Materials and methods

This multicenter retrospective study enrolled 347 ALL patients. Following automated segmentation of the axial skeleton on CT, radiomics features were extracted from both CT and PET images. An optimal feature subset was selected through a four-step screening process. Subsequently, four predictive models were developed: logistic regression (LR), support vector machine (SVM), multilayer perceptron (MLP), and AutoGluon-Tabular (AG). Model performance was evaluated using receiver operating characteristic (ROC) curve analysis with corresponding area under the curve (AUC) and decision curve analysis (DCA). Additionally, Shapley Additive exPlanations (SHAP) value analysis was employed to enhance model interpretability and identify key predictive features.

Results

The study cohort included 347 patients for Task 1 and 261 patients for Task 2. For Task 1, the AG model demonstrated the highest discriminative performance, achieving AUC values of 0.929, 0.909, and 0.741 in the training, validation, and test sets, respectively. For Task 2, the MLP model exhibited robust performance, with AUC values of 0.902, 0.812, and 0.745 in the training, validation, and test sets, respectively.

Conclusion

Radiomics signatures derived from 18F-FDG PET/CT imaging exhibit strong discriminative capability in distinguishing MRD-positive from MRD-negative ALL patients, as well as in stratifying low-burden versus high-burden disease states, highlighting their promising potential for non-invasive clinical application in leukemia management.

Introduction

Acute lymphoblastic leukemia (ALL) is a malignant neoplastic disease characterized by the clonal proliferation of abnormal B-cell or T-cell lineage lymphocytes within the bone marrow. Over the past several decades, clinical outcomes have significantly improved due to advances in therapeutic approaches, including chemotherapy, targeted therapy, and hematopoietic stem cell transplantation.

Nevertheless, relapsed or refractory (r/r) ALL is associated with a poor prognosis in adults and continues to be the leading cause of cancer-related mortality in children[1–3]. For these patients, the presence of non-central nervous system extramedullary disease (non-CNS EMD) and persistently positive minimal residual disease (MRD) are key factors linked to treatment failure and adverse clinical outcomes[4, 5].

Isolated bone marrow relapse is the most frequently observed pattern, accounting for 50% to 60% of relapsed cases[6]. Serial bone marrow aspirations remain the gold standard for monitoring intramedullary disease status, often necessitating multiple invasive and uncomfortable procedures throughout their treatment course. Moreover, the results may be influenced by technical variability and sampling errors, limiting the reliability of this method.

^{18}F-FDG PET/CT has been demonstrated to be effective in the detection of extramedullary disease involvement[7–10]. Additionally, evidence suggests that increased FDG uptake in the bone marrow on PET/CT may reflect the metabolic activity of leukemic cells[11]. However, research regarding the diagnostic value of ^{18}F-FDG PET/CT for assessing bone marrow involvement remains limited. Currently, bone marrow involvement on FDG PET/CT is primarily evaluated through visual assessment[12–14], which relies on the physician's experience and is associated with relatively low diagnostic accuracy. Li et al. developed a radiomics-based machine learning model using ^{18}F-FDG PET/CT to detect bone marrow involvement in patients with suspected relapsed acute leukemia. This model demonstrated significantly better diagnostic performance in terms of sensitivity, specificity, and accuracy compared to visual analysis (87.5%, 89.5%, 88.6% vs. 62.5%, 73.7%, 68.6%)[15]. Nevertheless, this study was limited by a small sample size and the absence of external validation. Moreover, patients with r/r ALL and high tumor burden are at increased risk of developing tumor lysis syndrome, a potentially life-threatening complication with a reported mortality rate of up to 20%, necessitating close clinical monitoring[16].

Therefore, this study aimed to: (1) develop and validate a novel machine learning model for the identification of bone marrow involvement, and (2) differentiate levels of bone marrow tumor burden in patients with r/r ALL.

Materials and methods

Patients

This retrospective study was approved by all relevant institutional review boards, and written informed consent was waived due to the retrospective nature. This multicenter study enrolled patients from three independent hospitals between August 2019 and March 2025. They were structured into two distinct cohorts. Cohort 1 included participants recruited from Beijing Gobroad Boren Hospital (Center A), who were selected through stratified random sampling and subsequently divided into a training set and a validation set at a ratio of 7:3. Cohort 2 served as the external test set and comprised patients from XXX Hospital (Center B) and XXX Hospital (Center C). The training set was utilized to develop the model, the

validation set to optimize hyperparameters and perform preliminary performance evaluations, and the test set to assess the model's generalization ability.

All patients were diagnosed with r/r ALL by attending hematologists[17]. Clinical information was extracted from the electronic medical records, and all routine blood test parameters were collected within 3 days before or after ^{18}F-FDG PET/CT scanning. Bone marrow aspiration was performed at the posterior superior iliac spine, and the obtained specimen underwent smear examination, morphological assessment, and minimal residual disease (MRD) testing. The time interval between bone marrow aspiration and ^{18}F-FDG PET/CT imaging was restricted to within one week to ensure temporal consistency of the findings. Additionally, patients did not receive granulocyte colony-stimulating factor (G-CSF) for at least two weeks before PET/CT imaging. Detailed inclusion and exclusion criteria are summarized in **Figure 1**.

^{18}F-FDG PET/CT acquisition

Patients needed to fast for at least 6 hours before the PET/CT scan, and their blood glucose levels needed to be less than 11.1 mmol/L. PET/CT acquisition was initiated approximately one hour after tracer injection, typically from the base of the skull to the mid-thigh. The additional scan was performed based on diagnostic requirements. PET/CT imaging for the training cohort and validation set was acquired using a uMI Vista scanner (reconstruction method: ordered-subsets expectation maximization (OSEM), 2 iterations and 20 subsets, United Imaging, China). The external test set was performed on PoleStar m660 PET/CT, PoleStar Flight PET/CT (reconstruction method: OSEM with time of flight (TOF), 2 iterations and 10 subsets, SinoUnion Healthcare Inc., China) and Biograph mCT PET/CT (reconstruction method: OSEM with TOF, 3 iterations and 21 subsets, Siemens, Germany). To protect patient privacy and enhance research credibility, the imaging data were anonymized.

Axial skeleton segmentation and feature extraction

Axial bone segmentation was automatically performed on CT images using XSCloud AI Service (SinoUnion Healthcare Inc.). Radiomic feature extraction was carried out with PyRadiomics (version 3.1.0), an open-source Python-based toolkit specifically designed for the standardized extraction of quantitative imaging features from medical images[18].

The extracted radiomic features were categorized into three main groups: shape-based features, first-order statistical features, and texture-based features. Texture features were derived from multiple matrices, including Gray Level Co-occurrence Matrix (GLCM), Gray Level Size Zone Matrix (GLSZM), Gray Level Run Length Matrix (GLRLM), Neighboring Gray Tone Difference Matrix (NGTDM), and Gray Level Dependence Matrix (GLDM).

Both first-order and texture features were extracted from the original images and the transformed images. The transformed images were generated using mathematical filters, including square, square root, logarithmic, exponential, gradient, and wavelet transformation (yielding 8 filtered images per

original). All CT and PET images were resampled to an isotropic voxel size of 1×1×1 mm^3in order to ensure spatial consistency. Intensity discretization was applied using a bin width of 25 Hounsfield Units (HU) for CT images and 0.3 Standardized Uptake Value (SUV) for PET images.

To minimize inter-scanner variability across the multiple imaging centers involved, the ComBat method was applied to adjust for batch effects[19]. Principal component analysis (PCA) was subsequently used to assess and visualize the effectiveness of batch effect correction achieved by ComBat. The overall workflow of the study is illustrated in **Figure 2**.

Radiomics feature selection

Prior to feature selection, all radiomic features were standardized using z-score normalization to ensure comparability across different scales. The feature selection process consisted of four sequential steps: (1) removal of zero-variance features to eliminate non-informative descriptors; (2) elimination of highly correlated features (absolute Pearson/Spearman correlation coefficient >0.8) through correlation analysis to reduce redundancy; (3) identification of significantly different features between MRD-positive and MRD-negative groups (Task 1), as well as between MRD-low and MRD-high groups (Task 2), using analysis of variance (ANOVA); and (4) application of the forward sequential feature selection (FSFS) method to determine the optimal feature subsets for each task independently. FSFS is an iterative greedy algorithm that incrementally constructs the optimal feature subset. The process begins with an empty set and sequentially adds the feature that yields the highest improvement in F1-score when evaluated with a k-nearest neighbors (KNN) classifier. This iterative selection continues until no further enhancement in model performance is observed.

Machine learning model establishment and validation

Using the optimal feature subsets identified for Task 1 and Task 2, respectively, four distinct classification models were developed: logistic regression, support vector machines (SVM), multilayer perceptron (MLP), and AutoGluon-Tabular (AG) automated machine learning framework[20]. The AG classifier utilized an advanced stacked ensemble architecture that integrated multiple base machine learning models in a strategic manner, including neural networks, gradient-boosted trees (LightGBM and CatBoost), Random Forests, Extremely Randomized Trees, and KNN algorithms. Through its multi-layer stacking framework, the system systematically optimized model combinations to improve predictive performance.

Model predictive performance was assessed using multiple evaluation metrics, including accuracy, sensitivity, precision, F1 score, receiver operating characteristic (ROC) curve analysis, area under the ROC curve (AUC), and decision curve analysis (DCA). In addition, model interpretability was evaluated using Shapley Additive exPlanations (SHAP)[21], a game-theoretic method that quantifies the importance of each feature across the entire cohort. This approach calculated the marginal contribution of each predictor variable by determining Shapley values, which reflected both the magnitude and direction (positive/negative) of each feature's impact on the model's predictions.

Statistical analysis

All statistical analyses were performed by Python (version 3.11.11) and R (version 4.3.0). The Kolmogorov-Smirnov test or the Shapiro-Wilk test was used to assess for normality. For continuous variables following a normal distribution, the Student's t-test for group comparisons and Pearson correlation for association evaluation. Non-normally distributed continuous variables were analyzed using nonparametric methods, specifically the Mann-Whitney U test for group comparisons and Spearman's rank correlation for association analysis. Chi-square test or Fisher's exact test was used for categorical variables. A two-sided *P*-value threshold of 0.05 was established as the criterion for statistical significance.

This retrospective study was approved by the institutional review board, and the requirement for informed consent was waived.

Results

Patient characteristics

A total of 347 patients (208 males and 139 females; mean age 20.5 years) were enrolled in this multicenter study, including 265 cases of B-cell acute lymphoblastic leukemia (B-ALL) and 82 cases of T-cell acute lymphoblastic leukemia (T-ALL). MRD assessment in Task 1 identified 261 MRD-positive cases (75.2%) versus 86 MRD-negative cases (24.8%). **Table 1** summarizes the differences in clinical characteristics between MRD-positive and MRD-negative cohorts across all participating centers. In Task 2, 161 patients (61.7%) were classified as having low MRD burden (<25%), whereas 100 patients (38.3%) were categorized as having high MRD burden (≥25%). Comparative analyses between these groups are presented in **Table 2**.

Feature extraction and selection

A total of 2632 radiomic features were extracted, with 1316 features derived from CT images and 1316 from PET images, respectively. These included 14 shape-based features, 252 first-order statistical features (18 features × 14 image types, derived from both original and transformed images), and 1050 texture features (75 features × 14 image types, 24 from GLCM, 16 from GLRLM, 16 from GLSZM, 5 from NGTDM, and 14 from GLDM). **Figure S1** illustrates the effectiveness of ComBat harmonization by displaying the distributions of multi-center data before and after batch effect correction.

Initially, zero-variance features were removed. Following the elimination of highly correlated features (absolute Pearson's correlation coefficient $|r| > 0.8$), a total of 265 features were retained for subsequent analysis.

Subsequent ANOVA-based feature selection retained 30 features for Task 1 and 50 features for Task 2, respectively. FSFS then identified optimal subsets consisting of 15 features for Task 1 and 10 features

for Task 2 (**Table S1 and S2**). The correlation matrices of these final feature subsets are shown in **Figure 3**.

Model construction and performance

Performance of the radiomic signature for distinguishing MRD-positive and MRD-negative

The AG classifier exhibited superior discriminative performance, with AUC values of 0.929 (95% CI: 0.889-0.963) on the training set, 0.909 (95% CI: 0.818-0.978) on the validation set, and 0.741 (95% CI: 0.615-0.851) on the test set, significantly outperforming other models (**Figure 4**). Further analysis identified that the AG classifier achieved higher accuracy, precision, and F1-score across both training and validation sets compared to the three benchmark models (**Table 3**). However, this performance advantage was not consistently observed in the external test set. Decision curve analysis (DCA, **Figure S2**) showed that the AG classifier provided a greater overall net benefit compared to the comparator models; however, its clinical utility was reduced in external validation, particularly at low-risk thresholds, where it exhibited relatively lower performance.

Performance of radiomic signature for distinguishing MRD-low and MRD-high

Although the AG model achieved near-perfect classification performance on the training set (AUC = 0.966, 95% CI: 0.935-0.988) and maintained acceptable performance on the validation set (AUC = 0.833, 95% CI: 0.729-0.925), its performance significantly declined on the external test set, yielding an AUC of 0.545 (95% CI: 0.362-0.715). In contrast, the MLP model showed more consistent and balanced performance across all datasets, with AUC values of 0.902 (95% CI: 0.845-0.951) on the training set and 0.812 (95% CI: 0.696-0.910) on the validation set. Notably, the MLP model exhibited the best generalization capability in the external validation set, achieving the highest AUC of 0.745 (95% CI: 0.579-0.885) among all evaluated models (**Figure 5**).

The AG classifier exhibited superior performance in terms of accuracy, precision, and F1-score compared to the three benchmark models across both training and validation sets. However, the MLP model achieved the highest performance across all evaluation metrics in the external test set (**Table 4**). DCA (**Figure S3)** indicated that the AG classifier provided greater net benefits at low-risk thresholds relative to the three benchmark models in both training and validation sets. In contrast, the MLP model demonstrated superior net benefit performance in the external validation cohort.

Model interpretation for SHAP

Global Shapley values were calculated to quantify feature importance, using the AG model for Task 1 and the MLP classifier for Task 2 (**Figure 6**). The bar charts illustrated global feature interpretation by visualizing the mean absolute Shapley values across all radiomic features, reflecting their overall contribution to the model's decision-making process. Beeswarm plots displayed visually the impact of individual features on prediction probability, where red and blue colors represented high and low feature values, respectively, and the distribution along the x-axis indicated the magnitude of each feature's

effect. Heatmap plots presented patient-level Shapley values for individual radiomic features, offering local interpretability while simultaneously illustrating both the direction and magnitude of each feature's influence on model predictions.

The top four most influential features in Task 1 were: PET_exponential_ngtdm_Coarseness(F5),PET_waveletLHL_glcm_ClusterProminence(F9),PET_wavelet-LLL_gldm_DependenceNonUniformity(F13),and PET_square_glszm_LowGrayLevelZoneEmphasis (F3). Notably, higher values of F5 and F3 were associated with MRD-negative predictions, while increased values of the other features were linked to MRD-positive classifications.

In Task 2, the most influential predictive features were: PET_wavelet-HLH_glszm_ZoneEntropy (F9), PET_wavelet-LLH_glcm_MCC (F5), CT_wavelet-HLH_firstorder_Skewness(F2),and PET_gradient_glszm_SmallAreaLowGrayLevelEmphasis (F4). All four features exhibited positive directional effects, with higher values consistently corresponding to MRD-high classification.

Discussion

As an emerging non-invasive methodology, radiomics facilitates the systematic quantification of imaging features by converting radiological data into high-dimensional, mineable datasets[22, 23]. These radiomic features offer a quantitative assessment of tumor spatial heterogeneity through computational analysis of voxel intensity patterns. Accumulating evidence has shown that PET/CT-based radiomic features can serve as reliable biomarkers for predicting bone marrow infiltration in various malignancies, including neuroblastoma[24], lymphoma[25, 26], and multiple myeloma[27]. Moreover, radiomics has also demonstrated predictive value for MRD status in multiple myeloma[28], and mantle cell lymphoma[29].

Although ^{18}F-FDG PET/CT is not currently a standard component of the diagnostic workup for AL, accumulating evidence has demonstrated that its high sensitivity in detecting EMD outside the central nervous system[7]. Notably, preliminary studies suggested a potential correlation between bone marrow FDG uptake and the metabolic activity of leukemia cells[11]. Li et al. developed a radiomics-based machine learning model using ^{18}F-FDG PET/CT to identify bone marrow involvement in patients with suspected relapsed acute leukemia. Their results showed that the machine learning model achieved significantly higher diagnostic accuracy compared to conventional visual analysis (88.6% vs. 68.6%, p = 0.041). The authors concluded that wavelet-based radiomic analysis of ^{18}F-FDG PET/CT, when combined with machine learning algorithms, offered a robust and non-invasive diagnostic approach that effectively complemented and enhanced traditional visual interpretation. However, due to the limited sample size and the lack of follow-up studies, the clinical utility of ^{18}F-FDG PET/CT for assessing leukemic burden in bone marrow remains to be fully validated.

Our multicenter study developed two radiomics models: one for differentiating leukemia patients with and without bone marrow infiltration, and another for assessing the extent of infiltration. For Task 1, the AG classifier demonstrated superior performance, achieving AUC values of 0.929 on the training set,

0.909 on the validation set, and 0.741 on the test set, outperforming other models. The most influential feature was PET_exponential_ngtdm_Coarseness. The NGTDM-based Coarseness quantifies the inverse of spatial intensity variation, where higher values indicate lower local intensity variation and greater texture uniformity[30]. Previous studies have confirmed its ability to distinguish benign and malignant tumors[31, 32]. Our results demonstrated that the NGTDM-based Coarseness texture feature exhibited significant discriminative power, with higher values being predictive of MRD-negative status.

The wavelet transform enables multiscale decomposition of images into distinct frequency components, with low-frequency elements representing overall morphological features and high-frequency components capturing fine textural details. Texture features derived from this decomposition can effectively characterize heterogeneous frequency distributions within tumor volumes[33]. Several studies have established the clinical value of wavelet-derived features in tumor classification[34, 35]. In Task 2, our result showed that the MLP model demonstrated balanced performance across all cohorts, with AUC values of 0.902 on the training set, 0.812 on the validation set, and 0.745 on the test set. The most influential features were PET_wavelet-HLH_glszm_ZoneEntropy, and PET_wavelet-LLH_glcm_MCC, for which higher values were associated with MRD-high classification.

There were several limitations of our study. First, the retrospective design of this study may introduce potential biases, including information bias, selection bias, and unmeasured confounding factors that could not be fully controlled for in our analyses. Second, although both AG and MLP models achieved relatively high AUC values across all three datasets, DCA revealed limited net benefits for both models in the external test set. This limitation was particularly evident in the performance of AutoGluon-Tabular on Task 1, which demonstrated only marginal clinical value. This may be attributed to the heterogeneity of PET/CT images across different centers. Although we applied ComBat harmonization to mitigate batch effects, complete elimination could not be assured due to the inherent limitations. Third, significant challenges remained for integrating radiomics features into routine clinical practice. Therefore, substantial improvements in both the generalizability and interpretability of radiomics models were required before these research findings can be translated into clinically applicable decision-support tools. Despite these limitations, to the best of our knowledge, this is the largest-scale study to date investigating bone marrow involvement using ^{18}F-FDG PET/CT. Our findings provided valuable insights into the clinical application of PET/CT in ALL, particularly in demonstrating how radiomic analysis of PET images may enable reliable, non-invasive assessment of both the presence and extent of bone marrow infiltration.

Conclusion

^{18}F-FDG PET/CT radiomic features allow for the non-invasive differentiation of bone marrow infiltration status and assessment of tumor burden in patients with r/r ALL. This approach has the potential to reduce invasive procedures to some extent while maintaining satisfactory diagnostic accuracy.

Abbreviations

MRD = minimal residual disease, r/r ALL = relapsed/refractory acute lymphoblastic leukemia, ROC = receiver operating characteristic, AUC = area under the curve, DCA = decision curve analysis

Declarations

Funding

This study has received funding by grants from National Natural Science Foundation of China (grant No. 82272034) and Beijing Municipal Natural Science Foundation (grant No. 7232031).

Acknowledgements

The authors would like to thank the staff of the Department of Nuclear Medicine of Beijing Friendship Hospital of Capital Medical University, Beijing Gobroad Boren Hospital, and Beijing Fengtai You'anmen Hospital, for their assistance in carrying out this study.

References

1. Bhojwani D, Pui CH. Relapsed childhood acute lymphoblastic leukaemia. Lancet Oncol. 2013;14(6):e205-17.
2. Forman SJ, Rowe JM. The myth of the second remission of acute leukemia in the adult. Blood. 2013;121(7):1077-82.
3. Sellar RS, Rowntree C, Vora AJ, Furness CL, Goulden N, Mitchell C, et al. Relapse in teenage and young adult patients treated on a paediatric minimal residual disease stratified ALL treatment protocol is associated with a poor outcome: results from UKALL2003. Br J Haematol. 2018;181(4):515-22.
4. Buchmann S, Schrappe M, Baruchel A, Biondi A, Borowitz M, Campbell M, et al. Remission, treatment failure, and relapse in pediatric ALL: an international consensus of the Ponte-di-Legno Consortium. Blood. 2022;139(12):1785-93.
5. Zhao Z, Hu Y, Li J, Zhou Y, Zhang B, Deng S. Applications of PET in Diagnosis and Prognosis of Leukemia. Technol Cancer Res Treat. 2020;19:1533033820956993.
6. Cooper SL, Brown PA. Treatment of pediatric acute lymphoblastic leukemia. Pediatr Clin North Am. 2015;62(1):61-73.
7. Zhou WL, Wu HB, Wang LJ, Tian Y, Dong Y, Wang QS. Usefulness and pitfalls of F-18-FDG PET/CT for diagnosing extramedullary acute leukemia. Eur J Radiol. 2016;85(1):205-10.

8. Yao X, Wang H, Yang J. 18 F-FDG PET/CT Finding of Bilateral Breast Relapse in a Male Acute Lymphoblastic Leukemia Patient. Clin Nucl Med. 2025;50(3):248-9.
9. Cunningham I, Kohno B. 18 FDG-PET/CT: 21st century approach to leukemic tumors in 124 cases. Am J Hematol. 2016;91(4):379-84.
10. Holland EM, Yates B, Ling A, Yuan CM, Wang HW, Stetler-Stevenson M, et al. Characterization of extramedullary disease in B-ALL and response to CAR T-cell therapy. Blood Adv. 2022;6(7):2167-82.
11. Arimoto MK, Nakamoto Y, Nakatani K, Ishimori T, Yamashita K, Takaori-Kondo A, et al. Increased bone marrow uptake of 18F-FDG in leukemia patients: preliminary findings. Springerplus. 2015;4:521.
12. Su K, Nakamoto Y, Nakatani K, Kurihara K, Hayakawa N, Togashi K. Diffuse homogeneous bone marrow uptake of FDG in patients with acute lymphoblastic leukemia. Clin Nucl Med. 2013;38(1):e33-4.
13. Li Z, Li C, Chen B, Shi L, Gao F, Wang P, et al. FDG-PET/CT versus bone marrow biopsy in bone marrow involvement in newly diagnosed paediatric lymphoma: a systematic review and meta-analysis. J Orthop Surg Res. 2021;16(1):482.
14. Badr S, Kotb M, Elahmadawy MA, Moustafa H. Predictive Value of FDG PET/CT Versus Bone Marrow Biopsy in Pediatric Lymphoma. Clin Nucl Med. 2018;43(12):e428-e38.
15. Li H, Xu C, Xin B, Zheng C, Zhao Y, Hao K, et al. (18)F-FDG PET/CT Radiomic Analysis with Machine Learning for Identifying Bone Marrow Involvement in the Patients with Suspected Relapsed Acute Leukemia. Theranostics. 2019;9(16):4730-9.
16. Rajendran A, Bansal D, Marwaha RK, Singhi SC. Tumor lysis syndrome. Indian J Pediatr. 2013;80(1):50-4.
17. Barsan V, Li Y, Prabhu S, Baggott C, Nguyen K, Pacenta H, et al. Tisagenlecleucel utilisation and outcomes across refractory, first relapse and multiply relapsed B-cell acute lymphoblastic leukemia: a retrospective analysis of real-world patterns. EClinicalMedicine. 2023;65:102268.
18. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017;77(21):e104-e7.
19. Leithner D, Schöder H, Haug A, Vargas HA, Gibbs P, Häggström I, et al. Impact of ComBat Harmonization on PET Radiomics-Based Tissue Classification: A Dual-Center PET/MRI and PET/CT Study. J Nucl Med. 2022;63(10):1611-6.
20. Lin CY, Guo SM, Lien JJ, Lin WT, Liu YS, Lai CH, et al. Combined model integrating deep learning, radiomics, and clinical data to classify lung nodules at chest CT. Radiol Med. 2024;129(1):56-69.
21. Xiang H, Xiao Y, Li F, Li C, Liu L, Deng T, et al. Development and validation of an interpretable model integrating multimodal information for improving ovarian cancer diagnosis. Nat Commun. 2024;15(1):2681.
22. Hu Y, Zhao X, Zhang J, Han J, Dai M. Value of (18)F-FDG PET/CT radiomic features to distinguish solitary lung adenocarcinoma from tuberculosis. Eur J Nucl Med Mol Imaging. 2021;48(1):231-40.

23. Kiss FJ, Járó AI, Máthé D, Benedek C, Madaras V, Manno-Kovács A, et al. Predicting PD-L1 expression status in NSCLC using radiomic analysis of 18 F-FDG-PET/CT images. Eur J Nucl Med Mol Imaging. 2025.
24. Feng L, Yang X, Lu X, Kan Y, Wang C, Zhang H, et al. Diagnostic Value of (18)F-FDG PET/CT-Based Radiomics Nomogram in Bone Marrow Involvement of Pediatric Neuroblastoma. Acad Radiol. 2023;30(5):940-51.
25. Mayerhoefer ME, Riedl CC, Kumar A, Dogan A, Gibbs P, Weber M, et al. [18F]FDG-PET/CT Radiomics for Prediction of Bone Marrow Involvement in Mantle Cell Lymphoma: A Retrospective Study in 97 Patients. Cancers (Basel). 2020;12(5).
26. Zhu X, Lu D, Wu Y, Lu Y, He L, Deng Y, et al. An interpretable machine learning model for predicting bone marrow invasion in patients with lymphoma via (18)F-FDG PET/CT: a multicenter study. BMC Med Inform Decis Mak. 2025;25(1):264.
27. An S, Huang G, Yu X, Liu J, Chen Y. The added diagnostic value of 18 F-FDG PET/CT radiomic analysis in multiple myeloma patients with negative visual analysis. Nucl Med Commun. 2024;45(3):244-52.
28. Milara E, Gómez-Grande A, Tomás-Soler S, Seiffert AP, Alonso R, Gómez EJ, et al. Bone marrow segmentation and radiomics analysis of [(18)F]FDG PET/CT images for measurable residual disease assessment in multiple myeloma. Comput Methods Programs Biomed. 2022;225:107083.
29. Bodet-Milin C, Morvant C, Carlier T, Frecon G, Tournilhac O, Safar V, et al. Performance of baseline FDG-PET/CT radiomics for prediction of bone marrow minimal residual disease status in the LyMa-101 trial. Sci Rep. 2023;13(1):18177.
30. Zhang Y, Ko CC, Chen JH, Chang KT, Chen TY, Lim SW, et al. Radiomics Approach for Prediction of Recurrence in Non-Functioning Pituitary Macroadenomas. Front Oncol. 2020;10:590083.
31. Chen S, Harmon S, Perk T, Li X, Chen M, Li Y, et al. Using neighborhood gray tone difference matrix texture features on dual time point PET/CT images to differentiate malignant from benign FDG-avid solitary pulmonary nodules. Cancer Imaging. 2019;19(1):56.
32. Suga M, Nishii R, Miwa K, Kamitaka Y, Yamazaki K, Tamura K, et al. Differentiation between non-small cell lung cancer and radiation pneumonitis after carbon-ion radiotherapy by (18)F-FDG PET/CT texture analysis. Sci Rep. 2021;11(1):11509.
33. Xing H, Hao Z, Zhu W, Sun D, Ding J, Zhang H, et al. Preoperative prediction of pathological grade in pancreatic ductal adenocarcinoma based on (18)F-FDG PET/CT radiomics. EJNMMI Res. 2021;11(1):19.
34. Feng L, Qian L, Yang S, Ren Q, Zhang S, Qin H, et al. Prediction for Mitosis-Karyorrhexis Index Status of Pediatric Neuroblastoma via Machine Learning Based (18)F-FDG PET/CT Radiomics. Diagnostics (Basel). 2022;12(2).
35. Zhang L, Dong D, Li H, Tian J, Ouyang F, Mo X, et al. Development and validation of a magnetic resonance imaging-based model for the prediction of distant metastasis before initial treatment of nasopharyngeal carcinoma: A retrospective cohort study. EBioMedicine. 2019;40:327-35.

Tables

Tables 1 to 4 are available in the Supplementary Files section.

Figures

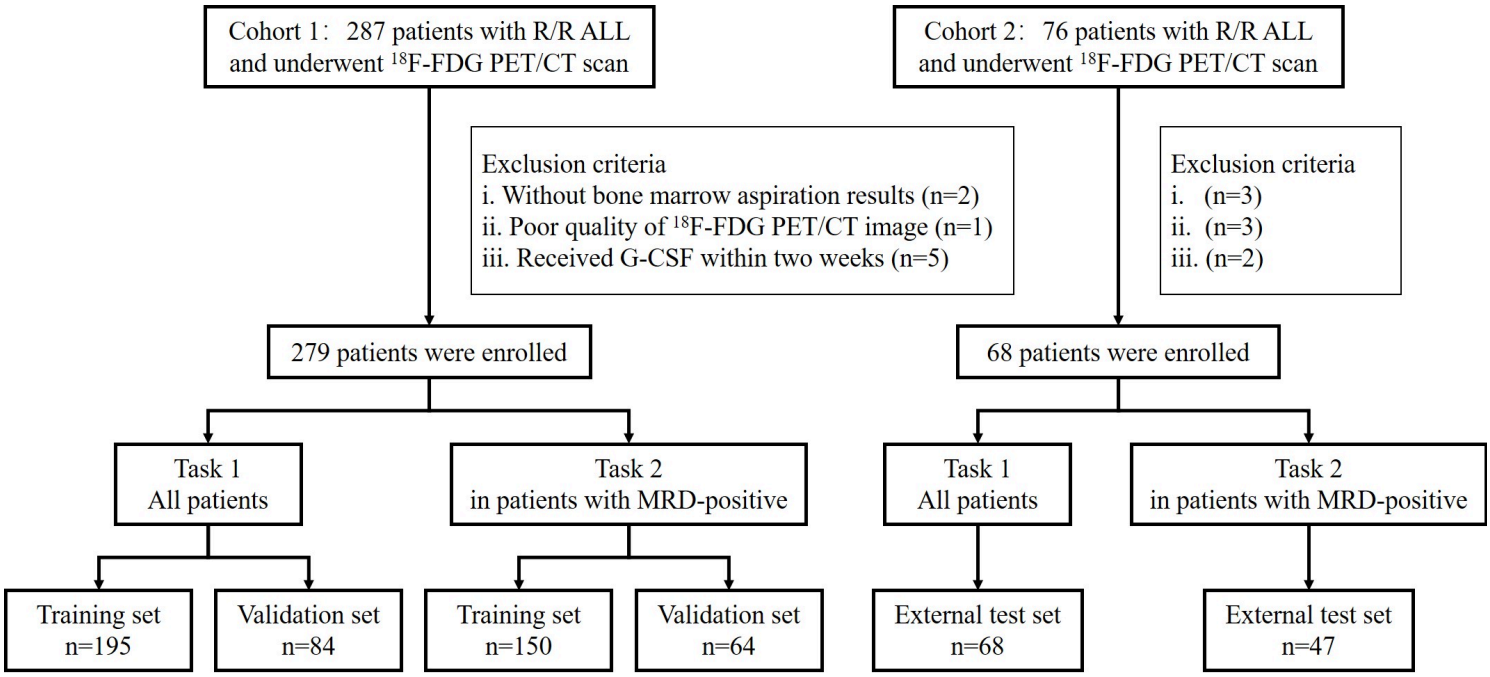


Figure 1

Flowchart of patient enrollment and exclusion criteria in the study.

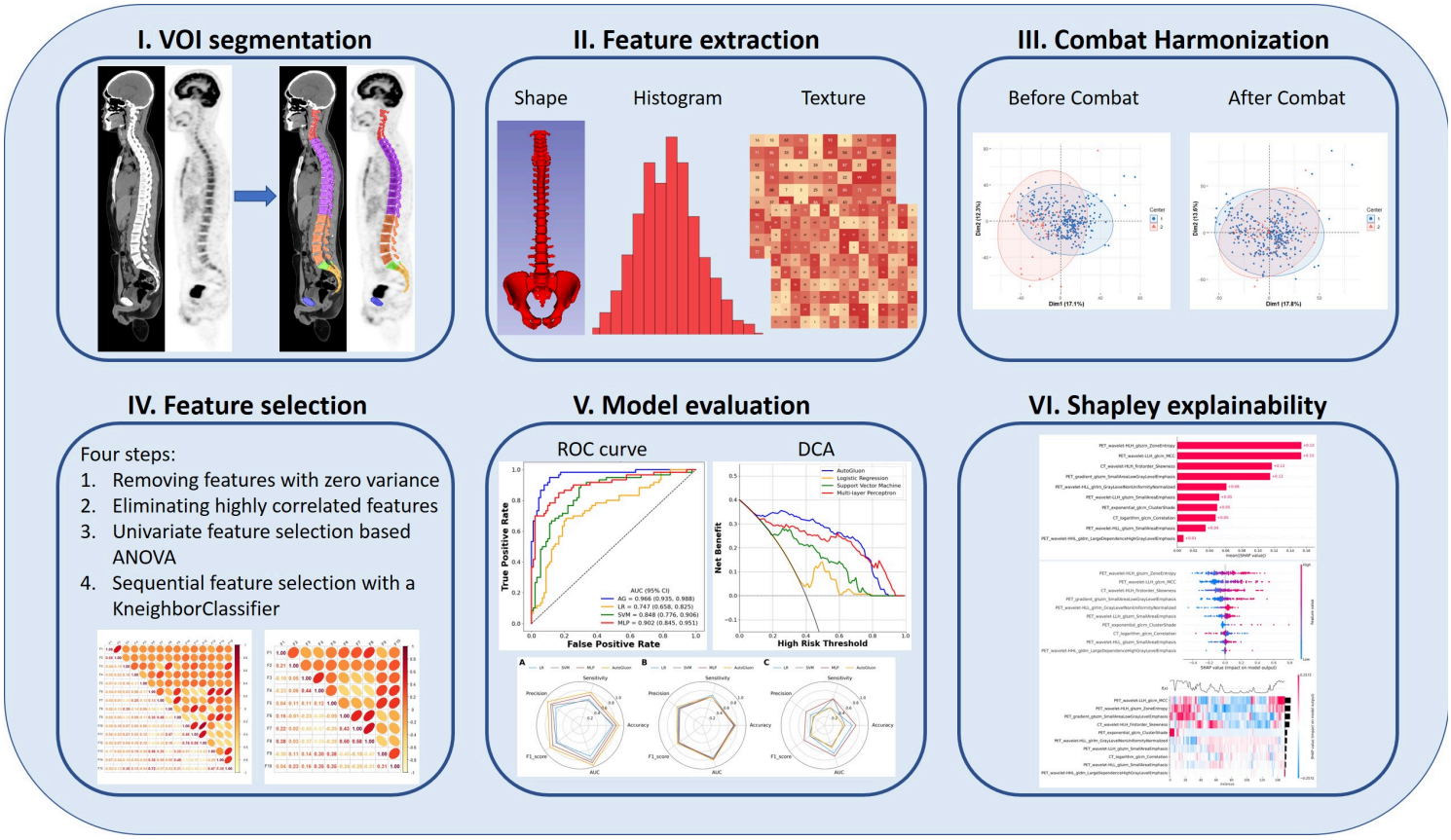


Figure 2

Workflow diagram of this study.

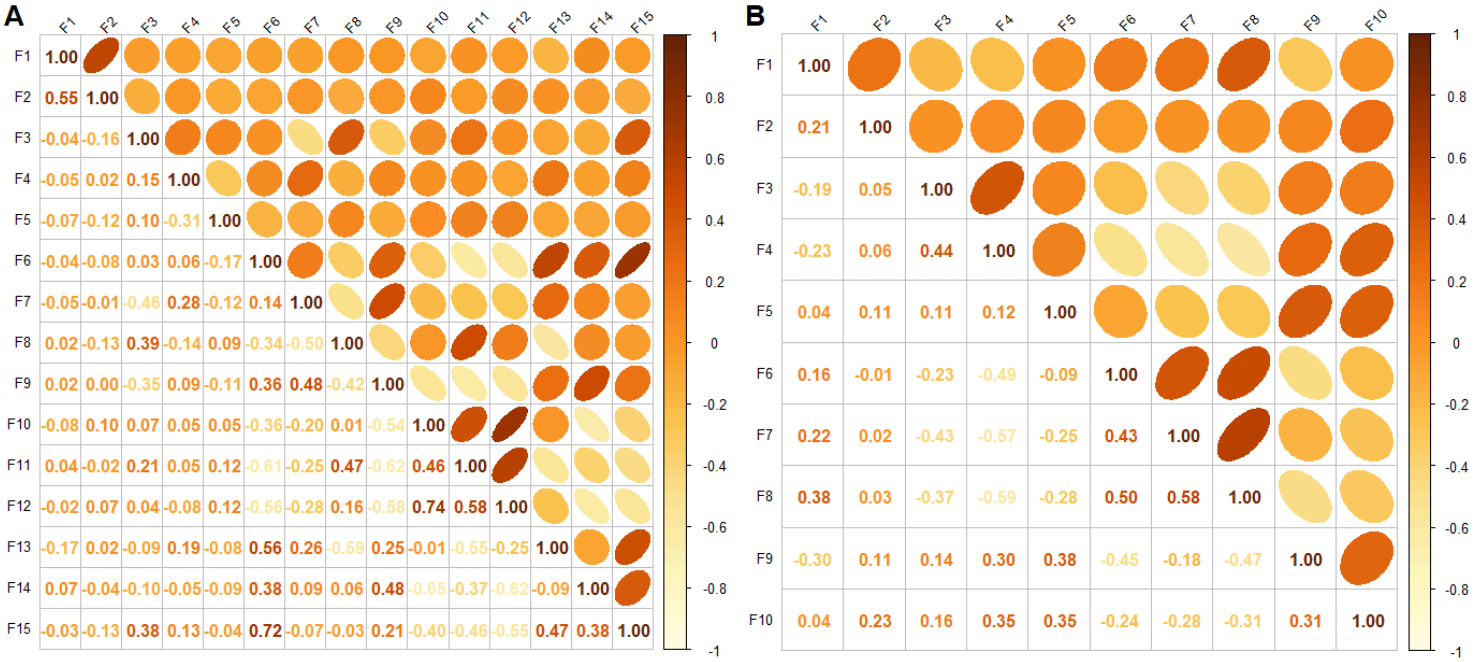


Figure 3

Heatmaps of correlation analysis. Correlation heatmap of the 15 radiomics-based features in Task 1(a), correlation heatmap of the 10 radiomics-based features in Task 2(b).

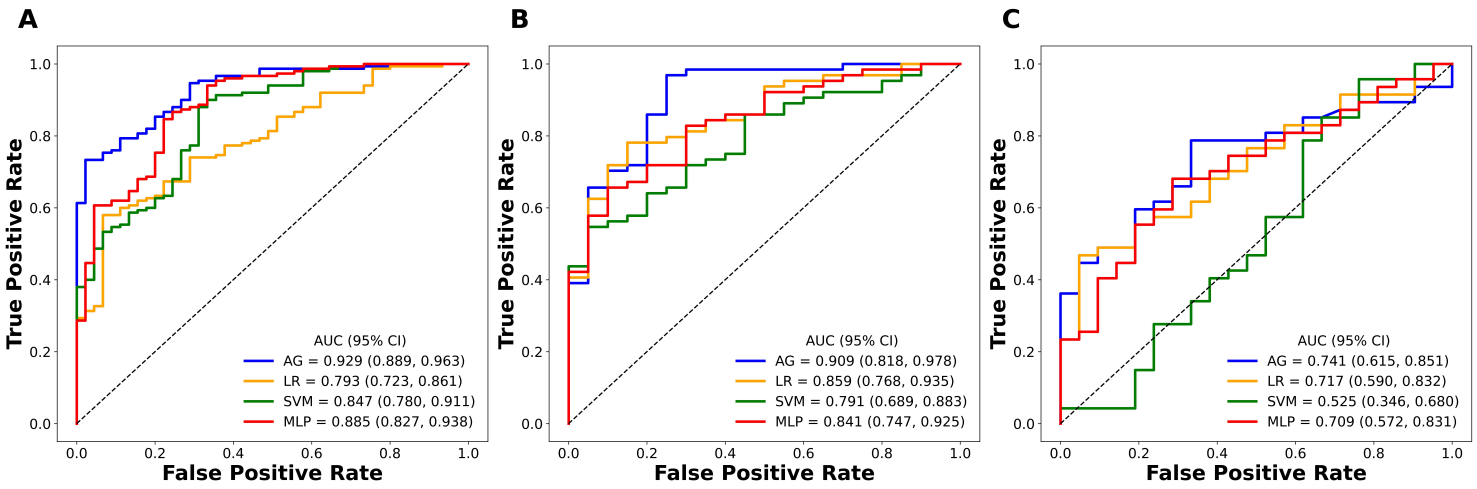


Figure 4

Receiver operating characteristic (ROC) curves in Task 1. ROC curves comparing the performance of different models on the training set(a), on the validation set models(b), and on the test set models(c).

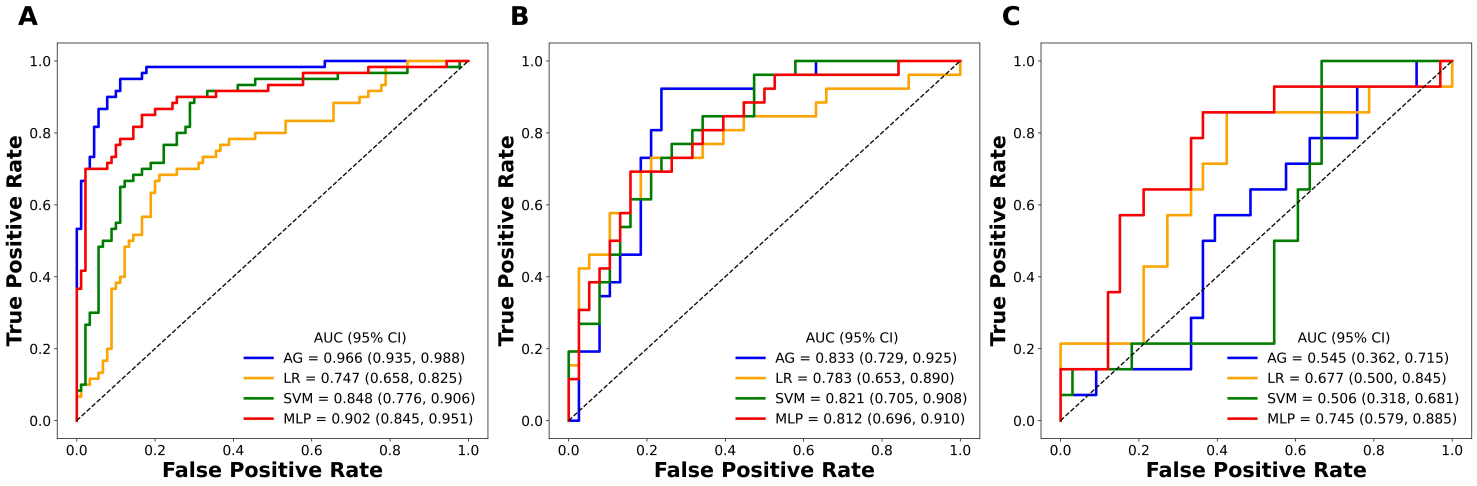


Figure 5

ROC curves in Task 2. ROC curves comparing the performance of different models on the training set(a), on the validation set models (b), and on the test set models(c).

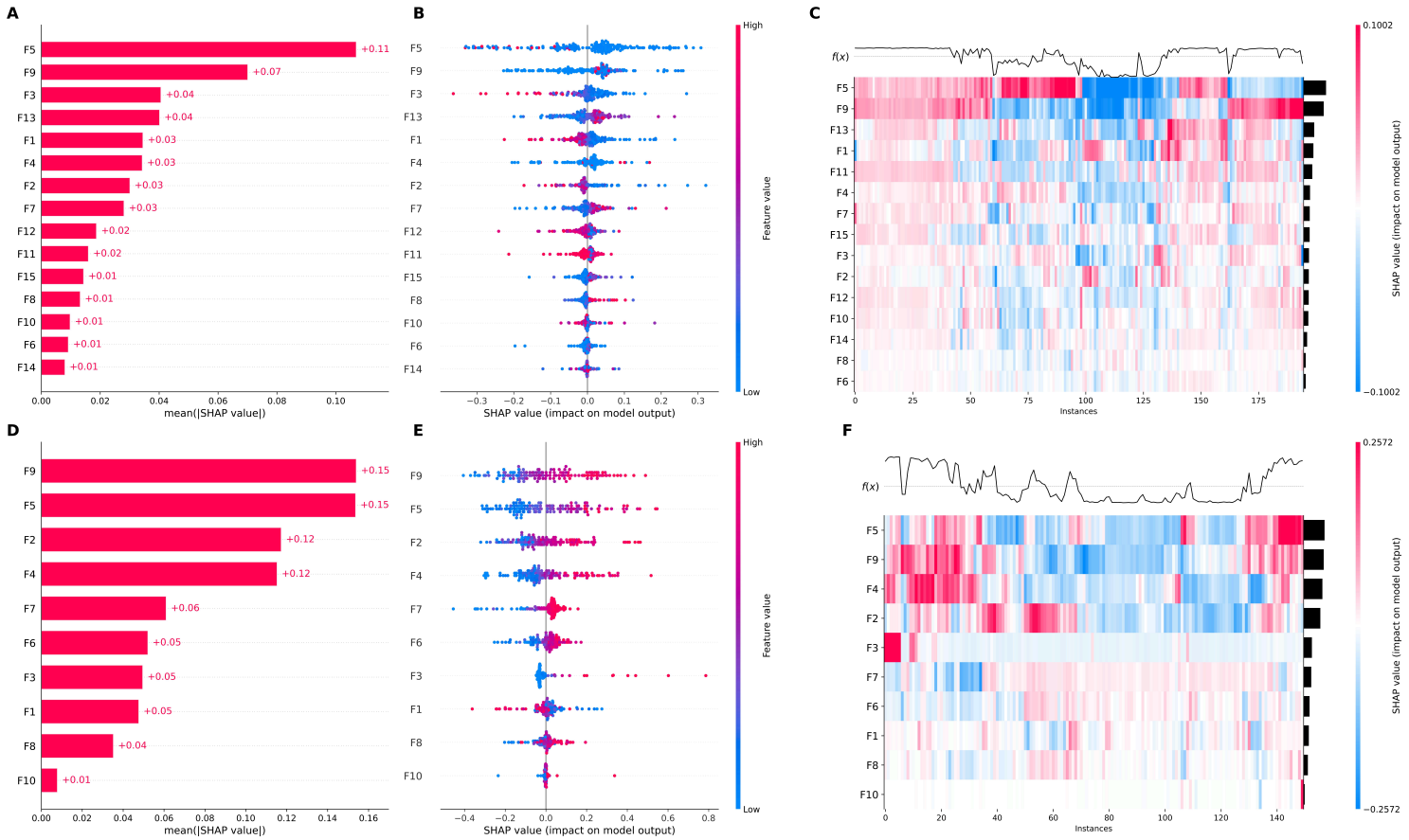


Figure 6

Shapley Additive exPlanations (SHAP) for Task 1 and Task 2. Bar chart(a), beeswarm plot(b), and heatmap plot (c).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.docx
- Table2.docx
- Table3.docx
- Table4.docx
- TableS1.docx
- TableS2.docx
- FigureS1Combat.png
- FigureS2task1DCA.jpg
- FigureS3task2DCA.jpg