Development of an AI model for predicting hypoxia status and prognosis in non-small cell lung cancer using multi-modal data

Lina Zhou[1#], Chenkai Mao[2#], Tingting Fu[3#], Xiao Ding[4,5], Luca Bertolaccini[6], Ao Liu[7], Junjun Zhang[2], Shicheng Li[2^]

[1]Health Management Center, The Second Affiliated Hospital of Soochow University, Suzhou, China; [2]Center for Cancer Diagnosis and Treatment, The Second Affiliated Hospital of Soochow University, Suzhou, China; [3]Department of Radiology, The Fourth Affiliated Hospital of Soochow University, Suzhou Dushu Lake Hospital, Suzhou, China; [4]State Key Laboratory of Common Mechanism Research for Major Diseases, Suzhou Institute of Systems Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, Suzhou, China; [5]Key Laboratory of Pathogen Infection Prevention and Control (Peking Union Medical College), Ministry of Education, Beijing, China; [6]Department of Thoracic Surgery, IEO, European Institute of Oncology IRCCS, Milan, Italy; [7]Department of Thoracic Surgery, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China

Contributions: (I) Conception and design: S Li; (II) Administrative support: L Zhou, C Mao, T Fu; (III) Provision of study materials or patients: J Zhang; (IV) Collection and assembly of data: X Ding; (V) Data analysis and interpretation: L Zhou, A Liu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Shicheng Li, MD; Junjun Zhang, MD. Center for Cancer Diagnosis and Treatment, The Second Affiliated Hospital of Soochow University, Sanxiang Road 1055, Suzhou 215123, China. Email: lishcheng@126.com; sudazhangjunjun@126.com; Ao Liu, MD. Department of Thoracic Surgery, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jingwu Weiqi Road 324, Jinan 250021, China. Email: liuaosmile@163.com.

Background: Prognosis prediction is crucial for non-small cell lung cancer (NSCLC) treatment planning. While tumor hypoxia significantly impacts patient outcomes, identifying hypoxic genomic markers remains challenging. This study sought to identify hypoxic computed tomography (CT) radiomic features and create an artificial intelligence (AI) model for NSCLC through the integration of multi-modal data.

Methods: In total, 452 NSCLC patients were enrolled in this study, including patients from The Second Affiliated Hospital of Soochow University (SC, n=112), The Cancer Genome Atlas (TCGA)-NSCLC dataset (n=74), the radiogenomics dataset (n=130), and the Gene Expression Omnibus (GEO) datasets (GSE19188: n=82, and GSE87340: n=54). Hypoxia status was classified using optimized cut-off values of hypoxia enrichment scores, which were calculated through single-sample gene set enrichment analysis (ssGSEA) of hypoxic genes. Radiomic features were extracted using three-dimensional (3D)-Slicer software. The least absolute shrinkage and selection operator (LASSO) algorithm was used to identify hypoxic CT radiomic features. A model named ssuBERT (semantic structured unit embedded in Bidirectional Encoder Representations from Transformers) was developed to analyze electronic health records (EHRs). An AI model for overall survival prediction was constructed by integrating CT radiomic features, ssuBERT features, and clinical data, and evaluated using five-fold cross-validation.

Results: Higher hypoxia levels were correlated with worse survival outcomes. Twenty-eight radiomic features showed significant discriminatory power in detecting hypoxia status with an area under the curve (AUC) of 0.8295. The ssuBERT model achieved a weighted accuracy of 0.945 in recognizing semantic structured units in EHRs. The EHR model exhibited superior predictive performance among the single-modal models with an AUC of 0.7662. However, the multi-modal AI model had the highest average AUC of 0.8449 and an F1 score of 0.7557.

^ ORCID: 0000-0002-5893-8822.

Conclusions: The AI model demonstrated potential in predicting NSCLC patient prognosis through multi-modal data integration, warranting further validation.

Keywords: Non-small cell lung cancer (NSCLC); hypoxia; radiomics; electronic health records (EHRs); prognostic model

Introduction

Lung cancer is the most common cancer in China, and non-small cell lung cancer (NSCLC) accounts for approximately 85% of all lung cancer cases (1). The accurate stratification of NSCLC patients according to survival predictions is essential for optimizing therapeutic interventions. Outcomes in lung cancer patients vary significantly due to multiple clinical determinants (2). Enhancing the overall clinical outcomes of patients is imperative. Consequently, there is an urgent need to develop robust prognostic models for predicting overall survival (OS) in NSCLC patients and informing clinical practice.

Hypoxia was initially associated with the "tumor microenvironment" of solid tumors. Tumors exhibit necrotic regions adjacent to areas with oxygen gradients influenced by proximity to blood vessels. While similar gradients are found in normal tissues, they show a pronounced and persistent decline in cancer (3). Moreover, cyclic hypoxia significantly activates the hypoxia-inducible factor (HIF) pathway, enhancing tumor cell survival under adverse conditions such as cytotoxic therapy (4). One recent study illustrated this overactivation using an analogy of tidal and wave cycles impacting the coast, with tides symbolizing hypoxia levels and waves representing high-frequency blood flow fluctuations. Overlaps in these cycles result in the overactivation of the HIF pathway (5). Research indicates that cyclic hypoxia exposure elevates cancer stem cell markers and promotes features associated with metastasis (6). While previous studies have explored the role of hypoxia in NSCLC development and outcomes (3,6), more extensive investigations are still needed.

Recent studies have advanced NSCLC prognosis prediction through linear models, with Liu *et al.* (7) incorporating hypoxia imaging biomarkers, Xu *et al.* (8) developing a liver metastasis nomogram, and Zhang *et al.* (9) creating a dynamic survival model. Meanwhile, artificial intelligence (AI) has demonstrated excellent performance in lung cancer prediction, with various approaches including machine learning, deep learning, and reinforcement learning being widely applied in this field. While these models demonstrate promising predictive capabilities, they face a main methodological challenge of limited integration of multi-modal markers like radiomics and free text, to enhance prediction accuracy and clinical applicability. In precision medicine, imaging-derived features can be correlated with pathological features, treatment responses, and survival outcomes, serving as crucial biomarkers that offer diagnostic, predictive, and prognostic insights (10). As predictive tools, this framework uses features extracted from images, including lesion volume, shape, and texture characteristics (11). The increasing accessibility of electronic health record (EHR) data, which efficiently includes extensive real-world populations, presents a cost-effective and timely alternative to conventional cohort studies (12).

Highlight box

Key findings
- Higher hypoxia levels are associated with worse survival in non-small cell lung cancer (NSCLC) patients.
- We developed an artificial intelligence (AI) model, which integrates ssuBERT (semantic structured unit embedded in Bidirectional Encoder Representations from Transformers), computed tomography (CT) radiomics, and clinical features. This multi-modal model outperformed single-modal approaches.

What is known, and what is new?
- Hypoxia in the tumor microenvironment negatively affects patient prognosis.
- This study identified 28 significant radiomic features related to hypoxia using CT imaging.

What is the implication, and what should change now?
- Integrating multi-modal data significantly enhanced the prediction of prognosis in NSCLC, providing a more accurate assessment.
- Additional research is needed to validate these findings in larger and more diverse cohorts, ensuring their robustness and generalizability.
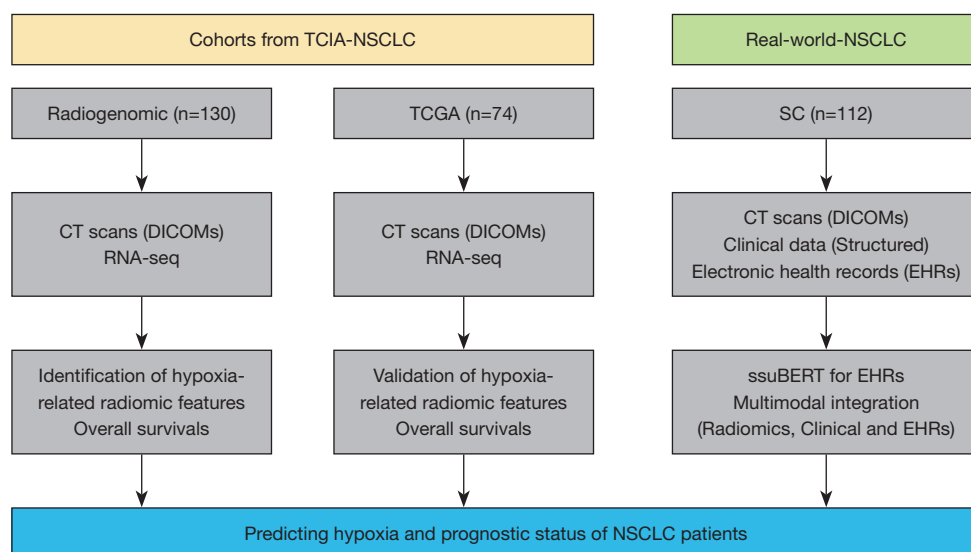
Figure 1 Patient enrollment of the study. TCIA, The Cancer Imaging Archive; NSCLC, non-small cell lung cancer; TCGA, The Cancer Genome Atlas; SC, The Second Affiliated Hospital of Soochow University; CT, computed tomography; DICOM, Digital Imaging and Communications in Medicine; ssuBERT, semantic structured unit embedded in Bidirectional Encoder Representations from Transformers; EHRs, electronic health records.

While numerous clinical variables are not directly captured in EHRs, these can be inferred from multiple data elements using machine-learning algorithms (13). Consequently, this study sought to develop robust models for assessing lung cancer prognosis using features derived from EHR and radiomic data. More specifically, this study aimed to predict hypoxia status through radiomic features and to construct a deep-learning prognostic model by integrating radiomic, EHR, and clinical data. We present this article in accordance with the TRIPOD reporting checklist (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-982/rc).

Methods

Patient enrollment

In this retrospective study, radiomics analyses were performed across the following cohorts: a real-world cohort of patients comprising the data of patients from The Second Affiliated Hospital of Soochow University (SC, n=112) and two cohorts comprising patient data from public databases, that is radiogenomics-NSCLC data (n=130) from The Cancer Imaging Archive (TCIA) (https://www.cancerimagingarchive. net/) and NSCLC data from The Cancer Genome Atlas (TCGA) (n=74) (14). *Figure 1* provides an overview of the

cohorts included in this study for the radiomics analysis.

To be eligible for inclusion in the SC cohort, patients had to meet the following criteria: (I) diagnosis of lung adenocarcinoma or lung squamous cell carcinoma; (II) undergone surgical or biopsy within four weeks of the computed tomography (CT) scan; and (III) lesion with a maximum diameter exceeding 1 cm to reduce the partial volume effect. Patients were excluded from the study if they met any of the following exclusion criteria: (I) a history of other malignancies; (II) incomplete or substandard case records; (III) poor-quality imaging data; (IV) lesions that were challenging to delineate on CT, such as those overlapping with adjacent structures or diffuse lesions; and (V) aged below 18 years. Ultimately, 112 patients from the SC were included in the study. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Medical Ethics Committee of The Second Affiliated Hospital of Soochow University (No. JD-HG-2023-63) and informed consent was taken from all the patients.

Study design

This retrospective study was conducted in three phases. In the first phase, radiomic features to predict hypoxia and
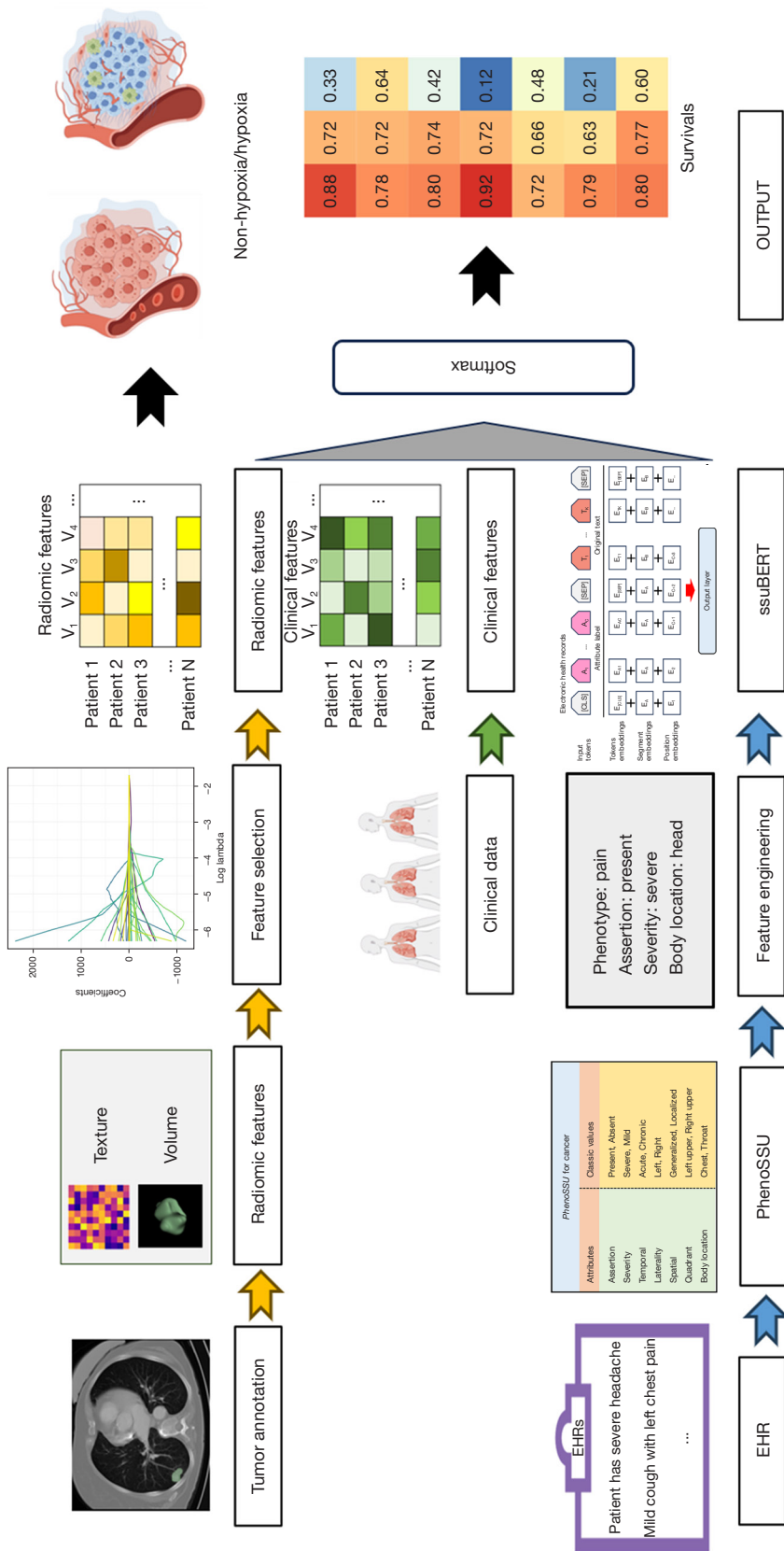
Figure 2 The structure of the multi-modal artificial intelligence model in the study. EHR, electronic health record; PhenoSSU, semantic structured unit of phenotypes; ssuBERT, semantic structured unit embedded in Bidirectional Encoder Representations from Transformers.

survival outcomes in NSCLC patients were identified. In the second phase, engineering features from EHRs were identified. In the third phase, these features were integrated using deep-learning techniques (*Figure 2*). Initially, we created a CT radiomic signature to forecast OS and hypoxia status. During this phase, the prognostic value of the hypoxia-related genes was assessed using cohorts from the Gene Expression Omnibus (GEO) datasets (GSE19188: n=82, and GSE87340: n=54). Next, we devised a method termed ssuBERT (semantic structured unit embedded in Bidirectional Encoder Representations from Transformers) to extract features from the EHRs. Finally, we constructed a deep-learning model by integrating CT radiomics, clinical features, and ssuBERT to predict NSCLC patient survival.

CT scan protocol

Patients of the SC cohort received CT scans using three different scanners: GE Revolution CT, GE CT 750 HD, or Philips Brilliance iCT. Scanning was performed with patients positioned supine and at full inspiration. Technical parameters for the scanning included: 150 kVp voltage, current ranging between 80 and 450 mA, detector collimation settings of either 64×0.625 mm or 128×0.625 mm, with the field of view set at 350 mm × 350 mm.

CT segmentation and structuring

CT data in Digital Imaging and Communications in Medicine (DICOM) format were imported into three-dimensional (3D)-Slicer software (https://www.slicer.org/) for all cohorts (Figure S1). Four experienced radiologists marked the primary tumor lesions as volumes of interest (VOIs), working without knowledge of patient information. The VOIs underwent preprocessing with resampling to $1\times1\times1$ mm^3 voxels and were rescaled to a range from –1,000 to 3,000 Hounsfield units (HU), using 10 HU as the fixed bin size. Following resampling and rescaling, radiomic features were extracted, including shape features, first-order statistics, and second-order statistics (four gray-level matrices). The reproducibility of manual image labeling was evaluated using the intraclass correlation coefficient (ICC) to assess intra-observer consistency. Radiomic features with an ICC value below 0.75 were deemed poor reproducibility and thus excluded. Finally, 127 radiomic features were retained for the subsequent analysis.

Identification of hypoxia-related radiomic features in NSCLC

The single-sample gene set enrichment analysis (ssGSEA) was used for calculating hypoxia enrichment scores. NSCLC patients were categorized into low- and high-hypoxia groups based on an optimal threshold identified using the Survminer package (version 0.4.9; https://rdocumentation.org/packages/survminer/versions/0.4.9). Hypoxia-related genes were sourced from the Msigdb database (https://www.gsea-msigdb.org/gsea/msigdb/, as detailed in Table S1). Kaplan-Meier curves with log-rank test were employed to compare OS between the different hypoxia groups of NSCLC patients. The least absolute shrinkage and selection operator (LASSO) algorithm, renowned for its effectiveness in managing high-dimensional collinear data, was applied to extract predictive features following data partitioning. We used a Python package to implement the LASSO process (alpha =1.0, max_iterance =1000, seed =12345), which was assessed using ROC (receiver operating characteristic) curves produced by the ROCR 1.1.0 package (15) in R.

The design of the PhenoSSU model for representing phenotype information in the EHRs of lung cancer patients

We previously developed a framework called PhenoSSU (semantic structured unit of phenotypes) to extract features from EHRs (16). To extract features from the EHRs of the NSCLC patients, we introduced a method called ssuBERT, which combines PhenoSSU and BERT (17). Initially, we refined the attributes of the PhenoSSU model to suit cancer patients better. The refined PhenoSSU model includes seven attributes. The structure of the PhenoSSU model is detailed in *Figure 3*. The PhenoSSU model allows for the structural representation of phenotype information from free text. Two Chinese annotators with medical expertise independently reviewed these medical records. The initial agreement between annotators, as measured by Cohen's kappa statistic, was 0.8771. A supervisor resolved any discrepancies in the annotations.

The workflow of recognizing PhenoSSU instances from EHRs

PhenoSSU instance recognition involves two primary subtasks: entity identification and attribute prediction. Our prior work established a robust framework for these tasks (18). The first subtask, entity identification, focused on detecting text spans corresponding to phenotype
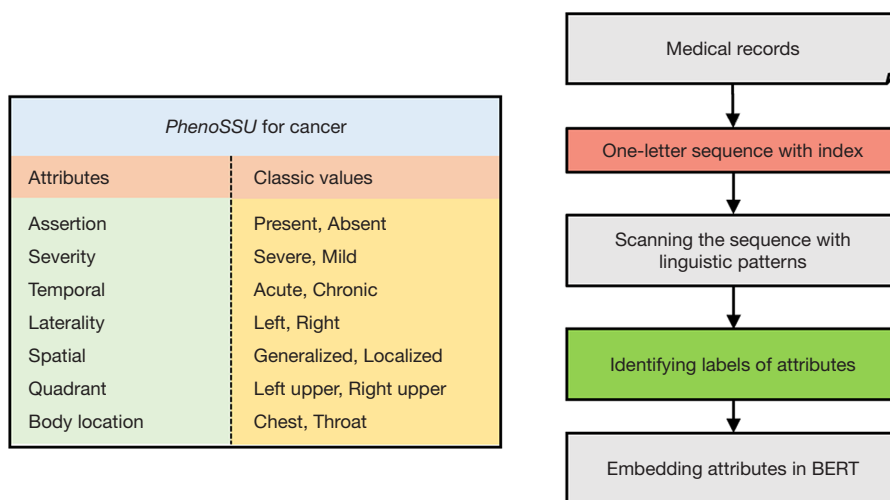
Figure 3 The design of the PhenoSSU model used for the EHR representation of lung cancer patients. PhenoSSU, semantic structured unit of phenotypes; EHR, electronic health record; BERT, Bidirectional Encoder Representations from Transformers.

and attribute entities. We employed two approaches for entity identification: (I) a deep learning-based method and (II) a dictionary-based method. The deep learning-based approach uses the advanced Bidirectional Encoder Representations from Transformers (BERT)-Bidirectional Long Short-Term Memory (BILSTM)-Conditional Random Field (CRF) model. The BERT model parameters were trained using the Kashgari package in Python (https://pypi.org/project/kashgari/). Convolutional neural network (CNN)-based models were also constructed for comparison. The dictionary-based method derives its phenotype knowledge base from Chinese translations of the International Classification of Diseases (ICD-11) (19). The second subtask involved predicting suitable values for attributes within the PhenoSSU model. For attribute prediction, we previously developed a pattern recognition-based method (18). Additionally, a conventional support vector machine (SVM) model was employed for the comparative analysis in this task.

Evaluation of the algorithm performance for recognizing PhenoSSU instances

The performance of PhenoSSU instance recognition was assessed using the metrics outlined in the Several-2015 Task 14: Analysis of Clinical Text (https://aclanthology.org//S15-2051). For entity recognition, the F1 score served as the primary evaluation metric. A predicted entity was deemed a true positive if it perfectly matched the gold-standard text span. Precision was defined as the ratio of correctly predicted entities to the total number identified by the algorithm. In contrast, recall was defined as the ratio of correctly predicted entities to those identified by annotators. The F1-score represents the harmonic mean of precision and recall. For attribute prediction, average accuracy and weighted average (WA) accuracy were used as the evaluation metrics. Notably, the WA accuracy accounts for the distribution of attribute values in the corpus, offering a more nuanced evaluation of infrequently occurring values.

ssuBERT: fusing PhenoSSU embedding into BERT

Drawing on BERT's sentence pair input configuration, we concatenated the attribute texts from PhenoSSU with the original document using a [SEP] token (separator token), applying distinct segment embeddings for the label texts and the document content. Document tokens are represented as T_i, with their respective embeddings denoted as E_{Ti}. Thus, T_K signifies the final token of the input document, where K represents the total number of words, and A_j represents the label text for the j_{th} class out of C total classes. Given that A_j may contain multiple subwords, E_{Aj}, the embedding of A_j, is computed by averaging the embeddings of all subwords within A_j. Consequently, the length of the label vector corresponds to C. Adhering to the methodology of the original BERT, a linear layer is used to prepare the input for
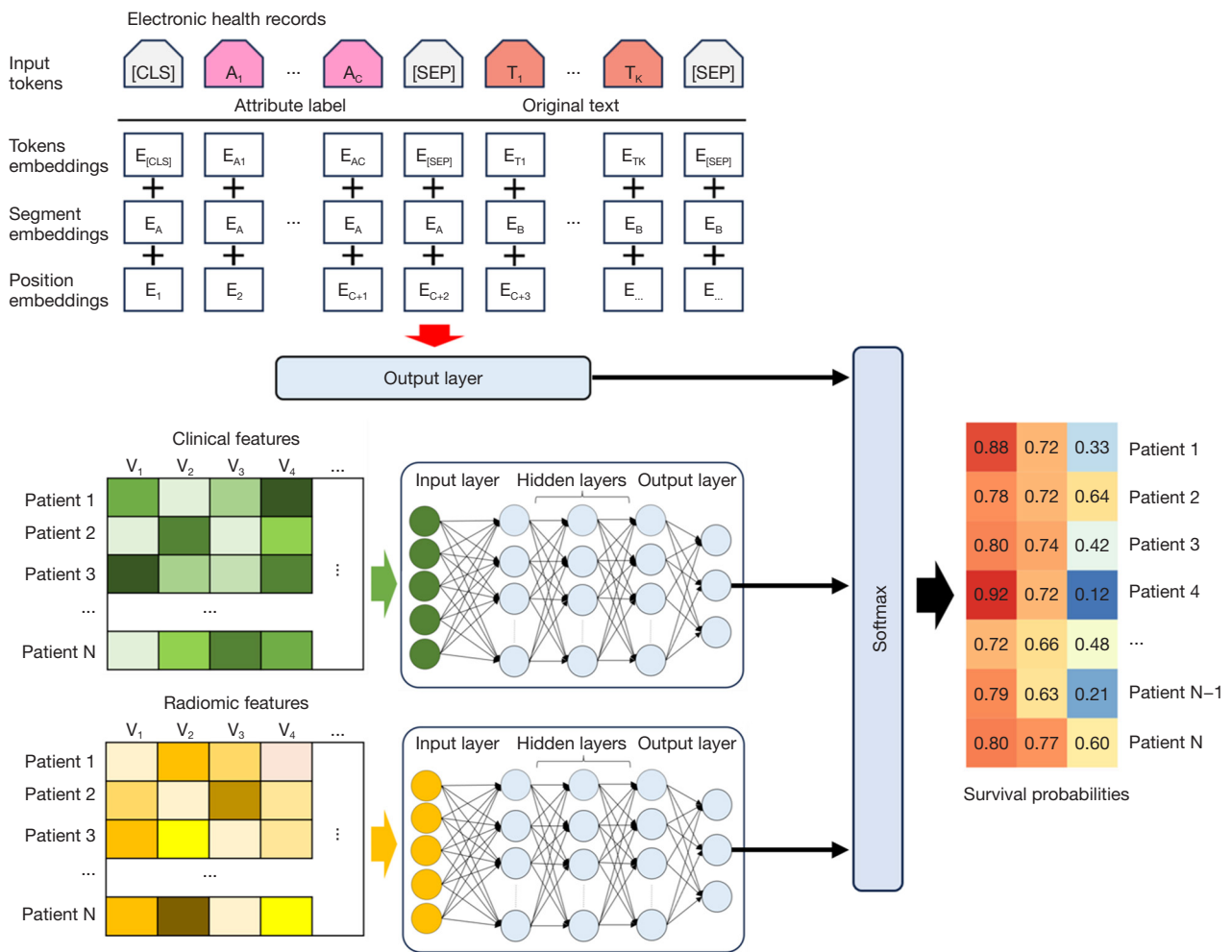
Transl Lung Cancer Res 2024;13(12):3642-3656 | https://dx.doi.org/10.21037/tlcr-24-982

Figure 4 The detailed structure of multi-modal data integration.

the SoftMax layer.

Model structure

We developed a multi-modal deep-learning model to predict clinical outcomes in NSCLC patients. Our objective was to predict patient survival using the following three input feature types: radiomic features, clinical characteristics, and embeddings from the linear layer of the ssuBERT model (see *Figure 4*). Deep neural networks (DNNs) were constructed for this purpose. Individual DNN models were trained on each feature set, and their performance was optimized by integrating them through a SoftMax layer. The DNN models were constructed and trained using TensorFlow (version 2.4.1) in Python. Separate

DNN models were trained on radiomic and clinical feature sets, each comprising three hidden layers using the ReLU (Rectified Linear Unit) activation function. The models were built with the Adam optimizer, a batch size of 128, a learning rate of 0.001, and a cross-entropy loss function. Post-training, the outputs from each model were merged and processed through a final Softmax layer to produce the ultimate survival prediction. The model was trained for 200 epochs using a GeForce RTX 4070 Ti GPU (Graphics Processing Unit). The models underwent five-fold cross-validation to assess their performance, which was evaluated using ROC curves with AUC, accuracy, F1 score, precision-recall AUC, and recall. Accuracy = (TP + TN)/(TP + TN + FP + FN). Recall = TP/(TP + FN). Precision= TP/(TP + FP). F1 = 2*(Precision * Recall)/(Precision + Recall). TP

Table 1 Characteristics of the enrolled cohorts

Features	TCGA (N=74)	Radiogenomic (N=130)	SC (N=112)
Age (years), median [25%, 75%]	68 [60–73]	69 [63–76]	69 [66–75]
Female, n (%)	55.40	43.85	48.20
Histology			
Adenocarcinoma	34	96	66
Squamous carcinoma	40	34	46
Stage			
I–II	50	–	70
III–IV	20	–	42
Not reported	4	130	0
Laterality			
Left	30	80	57
Right	44	50	55
Vital status			
Alive	42	85	49
Dead	32	45	63
OS (month), median [25%, 75%]	24.6 [4.4–48]	42.0 [14.9–59.4]	25.2 [9–35.6]
Smoking status			
Non-smoker	–	20	12
Current	–	25	36
Former	–	85	64

TCGA, The Cancer Genome Atlas; SC, The Second Affiliated Hospital of Soochow University; OS, overall survival.

(True Positive), TN (True Negative), FP (False Positive) and FN (False Negative).

Statistical analysis

The statistical analyses and visualizations were conducted with SPSS (Statistical Product and Service Solutions, version 25.0, IBM Corp., Armonk, NY, USA), GraphPad Prism (version 9.0), Python (version 3.7), Figdraw and R (version 3.5.3). The survival analysis and plotting were carried out using the survival package (version 3.5) and the survminer package (version 0.4.9) in R. A P value of less than 0.05 was considered statistically significant.

Results

Baseline information

Table 1 sets out the clinicopathological characteristics of the patients. The study included the radiomic data of 316 NSCLC patients, of whom 196 had lung adenocarcinoma and 120 had lung squamous carcinoma. Notably, the radiogenomic cohort had a higher proportion of lung adenocarcinoma patients (96/130) than the other two cohorts. The OS periods of the three cohorts were 24.6 (4.4–48) months for the TCGA cohort, 42.0 (14.9–59.4) months for the radiogenomic cohort, and 25.2 (9–35.6) months for the SC cohort.

Hypoxia is informative of the survival status of NSCLC

Using hypoxia gene sets, we first performed a ssGSEA to compute the hypoxia enrichment scores for samples from the TCGA, radiogenomic, and two GEO cohorts. Optimal cut-off values for hypoxia scores were calculated, enabling the classification of NSCLC patients from the radiogenomic, TCGA, GSE19188, and GSE87340 datasets into low and high-hypoxia groups (Figure S2A-S2D). Elevated hypoxia levels were correlated with worse survival outcomes across all cohorts (*Figure 5A-5D*).

Radiomic can predict hypoxia status of NSCLC

We used the LASSO regression model to select features from the radiomic data. A total of 28 radiomic features were extracted to develop the radiomic signature (*Figure 6A-6C*). Subsequently, the discriminative power of the radiomic features was assessed in the radiogenomic cohort, yielding an area under the curve (AUC) of 0.9235 (*Figure 6D*), and validated in the TCGA cohort, yielding an AUC of 0.8295 (*Figure 6E*). Building on these findings, we showed the prognostic significance of hypoxia genes and the 28 radiomic features in representing hypoxia status in NSCLC. These findings indicated that non-invasive CT methods hold promise for predicting survival outcomes in this cancer type.

The best strategy for recognizing PhenoSSU instances

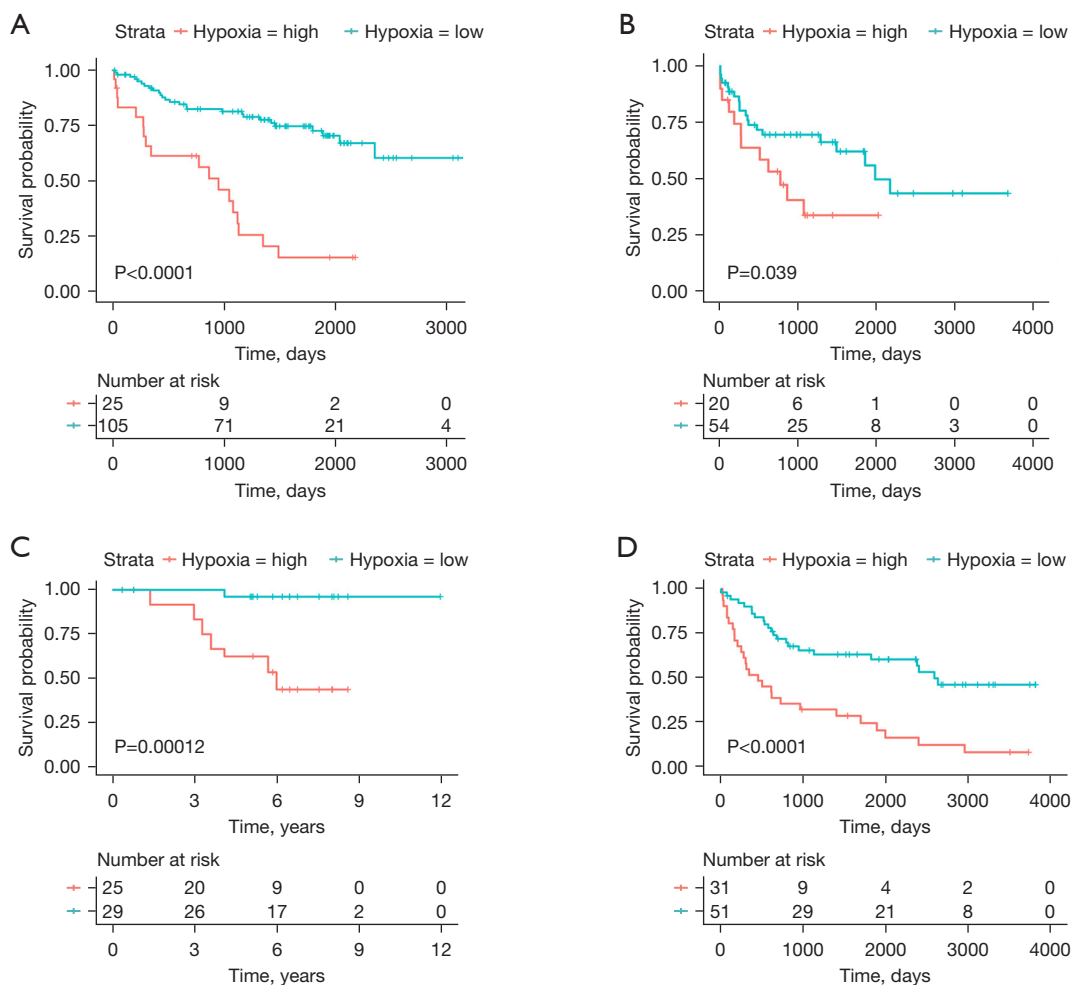To determine the optimal approach for identifying PhenoSSU instances, we evaluated and compared various

Figure 5 Survival analysis of the hypoxia scores. The survival curves of the samples from the radiogenomic (A), TCGA (B), GSE87340 (C), and GSE19188 (D) cohorts based on the best cut-off values, respectively. TCGA, The Cancer Genome Atlas.

techniques for the subtasks of entity recognition and attribute prediction. Specifically, the deep-learning approach using BERT-BILSTM-CRF yielded the highest performance in the entity recognition subtask, achieving an F1 score of 0.899 (*Table 2*). Conversely, the dictionary-based approach with ICD-11 had an F1-score of 0.789. For the attribute prediction subtask, the pattern recognition method, developed in our previous research (18), demonstrated superior performance, with a weighted accuracy of 0.945 (*Table 3*). Conversely, the SVM-based method attained a weighted accuracy of 0.731.

Using deep-leaning models to predict OS

DNN models were developed based on the different

predictors and a five-fold cross-validation analysis was conducted to predict the survival of patients with NSCLC (*Figure 7*). Among the single-modal models, the radiomic (*Figure 7A,7E*) and EHR (represented as ssuBERT; *Figure 7B,7F*) models showed superior predictive performance with AUC values of 0.7600 and 0.7662, and F1 scores of 0.6785 and 0.6733, respectively. The multi-modal model achieved the highest average AUC of 0.8449 and an F1 score of 0.7557 (*Figure 7D,7H*). By combining the modalities, the model benefited from the complementary strengths of each, leading to improved feature representation and enhanced predictive accuracy. This holistic approach allowed the multi-modal model to achieve higher performance metrics, as it leveraged a richer and more complete set of information for predicting patient survival.
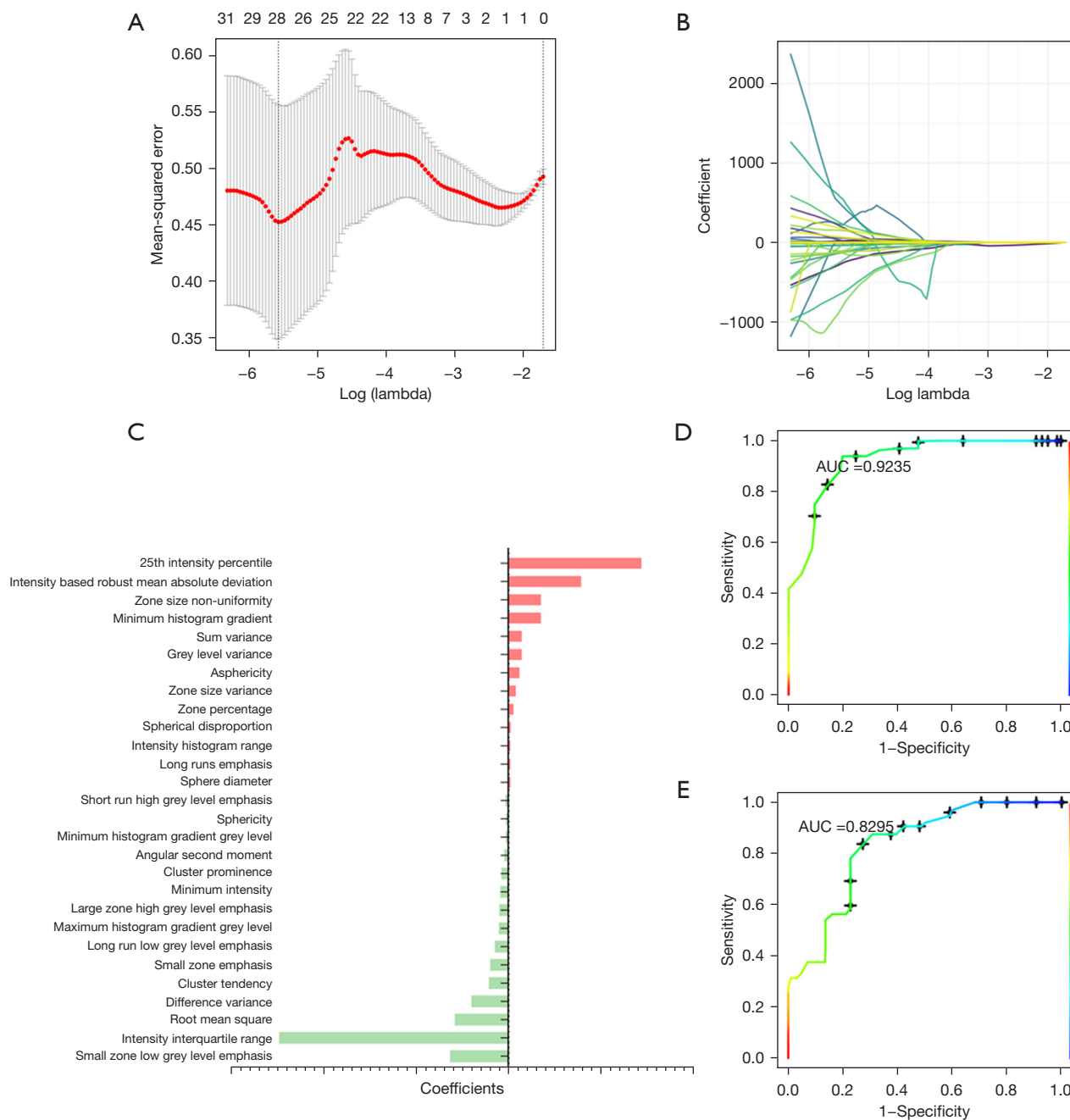
Figure 6 Hypoxia-related radiomic features derived from the LASSO model. (A,B) LASSO method for the screening of the radiomics features; (C) coefficients of the selected features, red indicates positive coefficients, while green indicates negative coefficients; (D,E) performance of the model. AUC, area under the curve; LASSO, least absolute shrinkage and selection operator.

Discussion

Lung cancer is the leading cause of cancer-related death worldwide, and poses a significant public health challenge (20). Hypoxia has emerged as a critical biomarker with promising

potential for clinical applications (3). The timely detection of hypoxic conditions is essential for selecting effective treatment strategies for NSCLC patients. However, conventional methods are invasive and face limitations, including limitations related to difficulties in sampling, constraints

Table 2 The performance of different methods in the subtasks of entity recognition

Method	AUC	Accuracy	Sensitivity	Specificity	F1 score
ICD-11	0.76	0.791	0.773	0.813	0.789
CNN	0.75	0.770	0.707	0.827	0.751
CNN-LSTM	0.79	0.811	0.799	0.826	0.811
BERT	0.82	0.837	0.801	0.881	0.833
BERT-BILSTM-CRF	0.85	0.903	0.879	0.920	0.899

AUC, area under the curve; ICD, International Classification of Diseases; CNN, convolutional neural network; LSTM, Long Short-Term Memory; BERT, Bidirectional Encoder Representations from Transformers; BERT-BILSTM-CRF, BERT-Bidirectional Long Short-Term Memory-Conditional Random Field.

Table 3 The performance of different methods in the subtasks of attribute prediction

Attribute	Pattern recognition (WA)	SVM (WA)
Assertion	0.971	0.802
Severity	0.932	0.722
Temporal pattern	0.909	0.729
Laterality pattern	0.821	0.673
Quadrant pattern	0.812	0.637
Spatial pattern	0.817	0.631
Body location	0.910	0.692
Total	0.945	0.731

WA, weighted accuracy; SVM, support vector machine.

related to tissue availability, and tumor heterogeneity. Conversely, non-invasive radiomic and EHR features offer the potential to predict clinical outcomes by uncovering relevant molecular information (21). This study established a deep-learning model designed to non-invasively predict the survival outcomes of NSCLC patients by integrating EHR, radiomic, and clinical feature data.

Hypoxia has been extensively investigated in cancer research with one recent study highlighting its potential clinical significance (22). Insufficient oxygen supply at the cellular, tissue, or organ level is frequently observed across various physiological and pathological states, and is regarded as a pivotal factor in carcinogenesis (23). Hypoxia serves as a critical sensor in key stages of cancer progression, including invasiveness, the acquisition of stem cell-like properties, stimulation of angiogenesis and lymph angiogenesis, and immune evasion, and also modulates radiotherapy sensitivity, cell survival, and resistance to apoptotic signals (24,25). Techniques such as immunohistochemistry commonly use HIF-1α and pimonidazole staining to study hypoxia (24,26). The expression of HIF-1α inducible markers and the synthesis of nitroimidazoles are influenced by varying oxygen levels, and HIFs can also be activated under physiological hypoxic conditions (27). Therefore, the accurate identification of hypoxic states may improve the prediction of patient prognosis and enhance clinical outcomes for lung cancer patients.

Numerous studies have examined the use of hypoxic gene profiles in predicting the survival of patients with various types of cancer, including gastric cancer (28), liver cancer (29), and lung cancer (30). However, the identification of genomic markers is often costly and complex. Consequently, this study focused on evaluating hypoxia status and predicting survival outcomes for NSCLC patients by combining radiomic features with EHRs.

Previous research has examined the predictive value of radiomic features and EHRs in lung cancer. For instance, one study identified clusters of radiomic features linked to lung cancer prognosis (31). Another study also established a relationship between CT radiomic features and OS in NSCLC patients (32). Additionally, a recent study explored the correlation between patient characteristics and clinical outcomes in NSCLC patients using EHRs (9). Another study assessed the efficacy of machine-learning algorithms in deriving a cohort of lung cancer patients from EHRs and in estimating their OS (33). Thus, the integration of radiomics and EHRs holds promise for enhancing survival prediction accuracy.

This study employed hypoxia-associated radiomics and EHRs to develop an advanced deep-learning model. By merging radiomic data with EHR information, the model delivers early and precise prognostic insights for patients with NSCLC. This capability will enable oncologists to customize treatment strategies more effectively, including modulating therapy intensity based on predicted outcomes. Further, the model can pinpoint high-risk patients who may benefit from intensified monitoring or enrollment in clinical trials, thereby potentially enhancing survival rates and the overall quality of life of patients. For instance, patients predicted to have a poor prognosis might be considered for experimental treatments or more frequent imaging, while those with a more favorable prognosis could avoid unnecessary overtreatment and its potential adverse effects. These capabilities illustrate how the model's predictions can
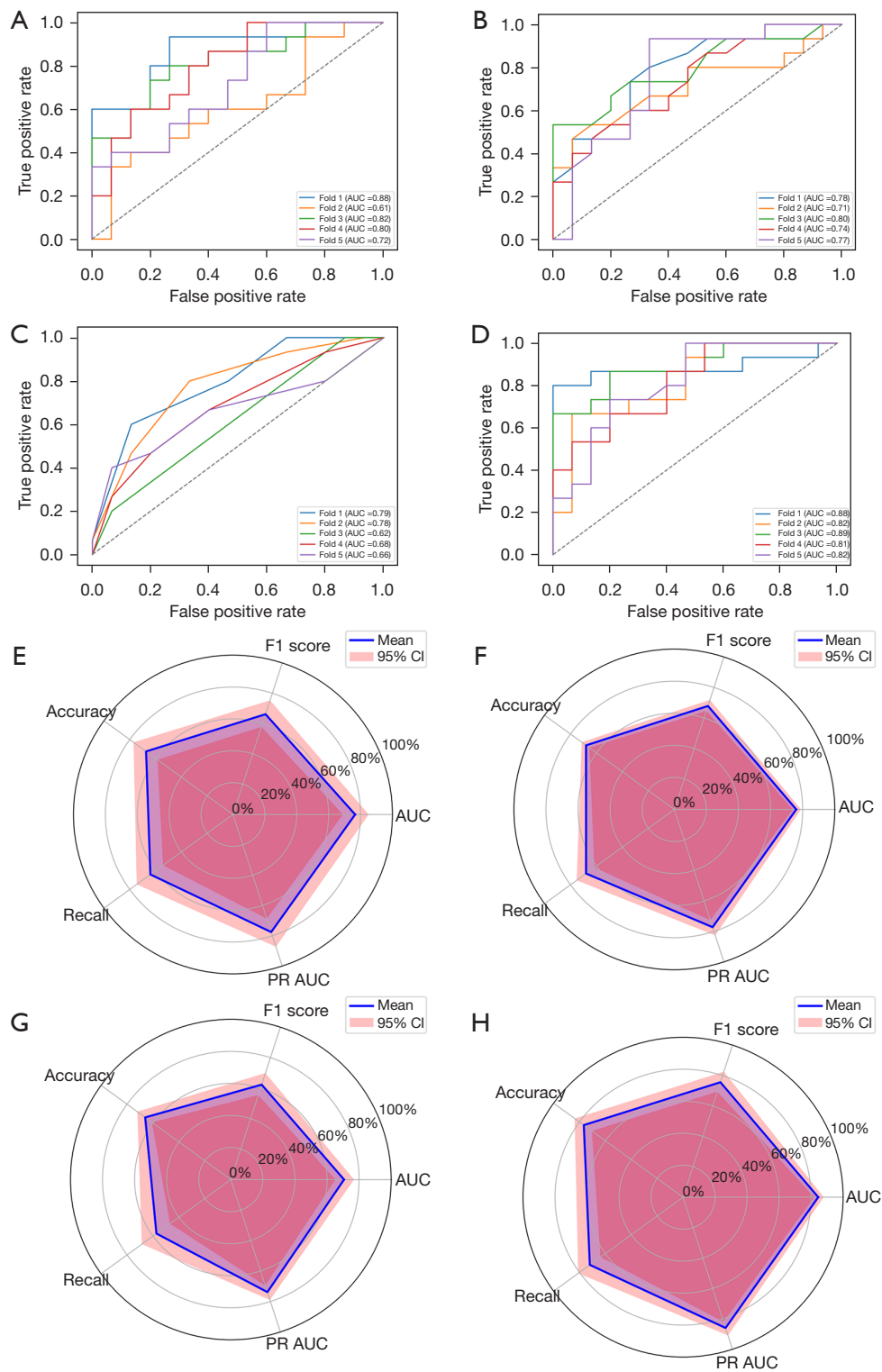
Figure 7 Performance of different models based on multi-modal data in the five-fold cross-validation. ROC curves of the radiomic (A), ssuBERT (B), clinical feature (C), and multi-modal (D) models. Radar plots of the radiomic (E), ssuBERT (F), clinical feature (G), and multi-modal (H) models, showing AUC, accuracy, F1 score, precision-recall AUC, and recall. AUC, area under the curve; CI, confidence interval; ROC, receiver operating characteristic; PR, precision-recall; ssuBERT, semantic structured unit embedded in Bidirectional Encoder Representations from Transformers.

advance personalized medicine, facilitating more informed clinical decisions and improved patient outcomes. Regarding external validation, while we acknowledge its importance, due to the current unavailability of a suitable external cohort with all required data modalities, we implemented a robust 5-fold cross-validation strategy. This approach helps mitigate overfitting, provides reliable estimates of model generalizability, and demonstrates consistent performance across folds, though we have acknowledged this limitation and highlighted it as a direction for future research.

This study had several limitations. First, despite our efforts to standardize image acquisition to three specific CT systems and pre-process the images prior to segmentation, discrepancies among imaging devices might have affected the results. Second, the semi-automatic tumor segmentation process proved to be labor-intensive for the radiologists. Third, the limited sample size of this study highlights the need for further prospective research to enhance the generalizability and robustness of the developed model. Finally, it is crucial to evaluate the prediction model on external datasets to address potential overfitting.

Conclusions

We introduced an innovative method for evaluating survival status through the integration of radiomic and EHRs. This methodology could assist clinicians to refine treatment strategies for patients with NSCLC.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-982/rc

Data Sharing Statement: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-982/dss

Peer Review File: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-982/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-24-982/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Medical Ethics Committee of The Second Affiliated Hospital of Soochow University (No. JD-HG-2023-63) and informed consent was taken from all the patients.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74:12-49.
2. Pei Q, Luo Y, Chen Y, et al. Artificial intelligence in clinical applications for lung cancer: diagnosis, treatment and prognosis. Clin Chem Lab Med 2022;60:1974-83.
3. Jing X, Yang F, Shao C, et al. Role of hypoxia in cancer therapy by regulating the tumor microenvironment. Mol Cancer 2019;18:157.

4. Liu X, Xie P, Hao N, et al. HIF-1-regulated expression of calreticulin promotes breast tumorigenesis and progression through Wnt/β-catenin pathway activation. Proc Natl Acad Sci U S A 2021;118:e2109144118.

5. Bader SB, Dewhirst MW, Hammond EM. Cyclic Hypoxia: An Update on Its Characteristics, Methods to Measure It and Biological Implications in Cancer. Cancers (Basel) 2020;13:23.

6. Hao S, Zhu X, Liu Z, et al. Chronic intermittent hypoxia promoted lung cancer stem cell-like properties via enhancing Bach1 expression. Respir Res 2021;22:58.

7. Liu M, Ma N, Ren C, et al. Hypoxia predicts favorable response to carbon ion radiotherapy in non-small cell lung cancer (NSCLC) defined by (18)F-FMISO positron emission tomography/computed tomography (PET/CT) imaging. Quant Imaging Med Surg 2024;14:3489-500.

8. Xu T, Liu X, Liu C, et al. Development and validation of a nomogram for predicting the overall survival in non-small cell lung cancer patients with liver metastasis. Transl Cancer Res 2023;12:3061-73.

9. Zhang Z, Xie S, Cai W, et al. A nomogram to predict the recurrence-free survival and analyze the utility of chemotherapy in stage IB non-small cell lung cancer. Transl Lung Cancer Res 2022;11:75-86.

10. Kummar S, Lu R. Using Radiomics in Cancer Management. JCO Precis Oncol 2024;8:e2400155.

11. Chen M, Copley SJ, Viola P, et al. Radiomics and artificial intelligence for precision medicine in lung cancer treatment. Semin Cancer Biol 2023;93:97-113.

12. Yuan Q, Cai T, Hong C, et al. Performance of a Machine Learning Algorithm Using Electronic Health Record Data to Identify and Estimate Survival in a Longitudinal Cohort of Patients With Lung Cancer. JAMA Netw Open 2021;4:e2114723.

13. Zhang Y, Cai T, Yu S, et al. High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (PheCAP). Nat Protoc 2019;14:3426-44.

14. Bakr S, Gevaert O, Echegaray S, et al. A radiogenomic dataset of non-small cell lung cancer. Sci Data 2018;5:180202.

15. Sing T, Sander O, Beerenwinkel N, et al. ROCR: visualizing classifier performance in R. Bioinformatics 2005;21:3940-1.

16. Deng L, Chen L, Yang T, et al. Constructing High-Fidelity Phenotype Knowledge Graphs for Infectious Diseases With a Fine-Grained Semantic Information Model: Development and Usability Study. J Med Internet Res 2021;23:e26892.

17. Niu H, Omitaomu OA, Langston MA, et al. EHR-BERT: A BERT-based model for effective anomaly detection in electronic health records. J Biomed Inform 2024;150:104605.

18. Li S, Deng L, Zhang X, et al. Deep Phenotyping of Chinese Electronic Health Records by Recognizing Linguistic Patterns of Phenotypic Narratives With a Sequence Motif Discovery Tool: Algorithm Development and Validation. J Med Internet Res 2022;24:e37213.

19. Harrison JE, Weber S, Jakob R, et al. ICD-11: an international classification of diseases for the twenty-first century. BMC Med Inform Decis Mak 2021;21:206.

20. Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: current status and future trends. Nat Rev Clin Oncol 2023;20:624-39.

21. Datta S, Bernstam EV, Roberts K. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. J Biomed Inform 2019;100:103301.

22. Wicks EE, Semenza GL. Hypoxia-inducible factors: cancer progression and clinical translation. J Clin Invest 2022;132:e159839.

23. Yang S, Hu C, Chen X, et al. Crosstalk between metabolism and cell death in tumorigenesis. Mol Cancer 2024;23:71.

24. Peng G, Liu Y. Hypoxia-inducible factors in cancer stem cells and inflammation. Trends Pharmacol Sci 2015;36:374-83.

25. Vito A, El-Sayes N, Mossman K. Hypoxia-Driven Immune Escape in the Tumor Microenvironment. Cells 2020;9:992.

26. Keith B, Simon MC. Hypoxia-inducible factors, stem cells, and cancer. Cell 2007;129:465-72.

27. Ancel J, Perotin JM, Dewolf M, et al. Hypoxia in Lung Cancer Management: A Translational Approach. Cancers (Basel) 2021;13:3421.

28. Deng C, Deng G, Chu H, et al. Construction of a hypoxia-immune-related prognostic panel based on integrated single-cell and bulk RNA sequencing analyses in gastric cancer. Front Immunol 2023;14:1140328.

29. Bao MH, Wong CC. Hypoxia, Metabolic Reprogramming, and Drug Resistance in Liver Cancer. Cells 2021;10:1715.

30. Shi Y, Fan S, Wu M, et al. YTHDF1 links hypoxia adaptation and non-small cell lung cancer progression. Nat Commun 2019;10:4892.

31. Tunali I, Gillies RJ, Schabath MB. Application of Radiomics and Artificial Intelligence for Lung Cancer

Precision Medicine. Cold Spring Harb Perspect Med 2021;11:a039537.

32. Huang B, Sollee J, Luo YH, et al. Prediction of lung malignancy progression and survival with machine learning based on pre-treatment FDG-PET/CT. EBioMedicine

2022;82:104127.

33. Elmarakeby HA, Trukhanov PS, Arroyo VM, et al. Empirical evaluation of language modeling to ascertain cancer outcomes from clinical text reports. BMC Bioinformatics 2023;24:328.