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.

Interpretable machine learning for predicting bone metastasis at initial diagnosis in elderly patients with small cell lung cancer: dual-centre development and external validation in two independent cohorts

Dajiang Sun
Hebei North University
Yifei Zheng
Hebei North University
Zhihao Ma
Chinese PLA General Hospital
Jiayi Yu
Peking University Cancer Hospital
Haodong Chen
the Fourth Medical Center of PLA General Hospital
Bu Wang
The First Clinical Medical College of the First Affiliated Hospital of Hebei North University
Weibo Liu
the Fourth Medical Center of PLA General Hospital
Wenhao Hu
Luckyhao555@163.com
the Fourth Medical Center of PLA General Hospital

Research Article

Keywords: small cell lung cancer, bone metastasis, prediction model, machine learning, LightGBM

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11570772/v1

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

Additional Declarations: No competing interests reported.

Interpretable machine learning for predicting bone metastasis at initial diagnosis in elderly patients with small cell lung cancer: dual-centre development and external validation in two independent cohorts

Author information

Dajiang Sun[1+], Yifei Zheng[1+], Zhihao Ma[2+], Jiayi Yu[3], Haodong Chen[4], Bu Wang[5*], Weibo Liu[6*], Wenhao Hu[6*]

[1]Graduate School, Hebei North University, Zhangjiakou, Hebei, China.

[2]Department of Orthopaedics, Chinese PLA General Hospital (PLA Medical School), Haidian District, Beijing, China.

[3]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiation Oncology, Peking University Cancer Hospital Institute, Beijing, China.

[4]Department of Joint Surgery, the Fourth Medical Centre, Chinese PLA General Hospital, Beijing, China.

[5]The First Clinical Medical College of the First Affiliated Hospital of Hebei North University, Zhangjiakou, Hebei, China.

[6]Department of Orthopaedics, the Fourth Medical Centre, Chinese PLA General Hospital, Haidian District, Beijing, China.

***Corresponding author:**

Bu Wang: The First Clinical Medical College of the First Affiliated Hospital of Hebei North University, Zhangjiakou, Hebei, China. email: 289974726@qq.com

Weibo Liu: Department of Orthopaedics, the Fourth Medical Centre, Chinese PLA General Hospital, Haidian District, Beijing, China. email: liuweibo2011@163.com

Wenhao Hu: Department of Orthopaedics, the Fourth Medical Centre, Chinese PLA General Hospital, Haidian District, Beijing, China. email: Luckyhao555@163.com

+:These authors contributed equally to this work.

Abstract

Background: Small-cell lung cancer (SCLC) is aggressive and often metastatic at diagnosis, and skeletal-related events such as bone pain, pathological fracture and spinal cord compression alter treatment decisions. Prediction tools for bone metastasis in elderly patients with SCLC remain scarce, and external validation and interpretability analyses are limited. We therefore developed and externally validated an interpretable machine learning model for bone metastasis at initial diagnosis in these patients, using routinely available clinical and staging information.

Methods: We retrospectively enrolled 1,387 consecutive patients aged ⩾60 years with SCLC treated at two hospitals (2010–2024), randomly divided 7:3 into training and internal validation sets. Two external cohorts were assembled: 128 patients from another hospital (2010–2024) and 19,561 patients from the SEER database (2010–2015). The outcome was bone metastasis at initial diagnosis. Boruta selected features from 16 candidate variables. Six models—logistic regression, decision tree, random forest, XGBoost, LightGBM and artificial neural network—were compared for discrimination and calibration, with decision curve analysis and SHAP interpretation.

Results: Boruta confirmed four features: N stage, SCLC stage (limited-stage or extensive-stage), liver metastasis and clinical stage. LightGBM was the most stable across cohorts. SHAP identified clinical stage as the largest contributor, with extensive-stage disease, concurrent liver metastasis and higher N stage increasing the predicted risk.

Conclusions: A LightGBM model based on four features available at initial diagnosis provides stable, interpretable risk stratification for bone metastasis in elderly patients with SCLC and may support clinical decision-making, although recalibration should precede use in other populations.

Keywords: small cell lung cancer; bone metastasis; prediction model; machine learning; LightGBM;

Background

Small cell lung cancer (SCLC) accounts for approximately 15% of lung cancers and is a rapidly proliferating, highly aggressive neuroendocrine tumour with a marked propensity for early metastasis [1]. Smoking is the predominant risk factor, and evidence that environmental exposures such as outdoor air pollution are carcinogenic continues to accumulate [2]. SCLC is typically diagnosed at an older age, and patients often have multiple comorbidities; age itself also influences the organ distribution of distant metastases and prognosis [3,4].

Most patients already have extensive-stage disease at initial diagnosis [1,5], and bone is one of the most common sites of distant metastasis in this population [6,7]; the prognostic impact of metastasis differs according to the organ involved [8,9]. Patients with bone metastasis are prone to skeletal-related events such as bone pain, pathological fracture and spinal cord compression, which frequently require additional radiotherapy, analgesia and bone-modifying agents and have a particularly pronounced effect on quality of life in elderly patients.

In initial staging, bone metastasis is detected mainly by whole-body bone scintigraphy or PET/CT. Current guidelines recommend comprehensive staging for SCLC but do not specify which patients require additional bone scintigraphy, let alone provide risk-based stratification [10]. Performing whole-body bone scintigraphy for every newly diagnosed patient is not cost-effective and entails radiation exposure, whereas selecting patients by clinical judgement alone risks missing asymptomatic bone metastasis. If individual risk of bone metastasis could

be estimated in advance from routine staging information, the allocation of imaging resources would be better grounded, and prediction models can play a role here [11].

Existing studies that predict bone metastasis in SCLC from SEER and other databases have relied mainly on nomograms or a single machine learning algorithm, with development and validation usually confined to one database, so evidence from independent external cohorts remains limited [12,13]. In recent years, interpretable machine learning has begun to be applied to metastasis risk prediction in SCLC in order to mitigate the difficulty of accepting black-box models in clinical practice [14]; however, work that focuses on elderly patients with SCLC while covering both hospital cohorts and a SEER population and that completes external validation is still lacking.

Against this background, the present study addressed three questions: whether usable risk stratification for bone metastasis at initial diagnosis can be achieved in elderly patients with SCLC using routine clinical and staging information alone; which of six commonly used algorithms (logistic regression, decision tree, random forest, XGBoost, LightGBM and artificial neural network) performs more stably across populations; and whether a model can deliver interpretable output and clear clinical net benefit while retaining discrimination. To this end, we developed and internally validated models using data from two hospitals, performed external validation in an independent Chinese hospital cohort and a SEER cohort, and comprehensively evaluated the representative model using calibration curves, the Brier score, decision curve analysis (DCA) and SHAP.

Methods

Study design and data sources

This was a retrospective, multi-cohort study reported in accordance with the TRIPOD+AI statement [15]. The development cohort comprised patients with SCLC treated consecutively

at the Chinese PLA General Hospital and Peking University Cancer Hospital between 2010 and 2024 (8,596 patients screened) and was used to derive the training and internal validation sets. External validation cohort 1 comprised patients with SCLC treated at the First Affiliated Hospital of Hebei North University between 2010 and 2024 (1,506 patients screened). External validation cohort 2 comprised elderly patients with SCLC diagnosed between 2010 and 2015 in the United States Surveillance, Epidemiology, and End Results (SEER) database (20,124 patients screened), with data extracted using SEER*Stat version 9.0.43. The patient selection process is shown in Figure 1. Detailed technical specifications of the cohort assembly, variable coding, missing-data handling, feature selection, model hyperparameters, model evaluation and interpretability analyses are provided in Additional file 10: Supplementary Methods.

Study population and outcome

Eligibility criteria for the two-hospital cohort were pathologically confirmed SCLC, age 60 years or older, and complete records of clinical stage, SCLC stage (limited-stage or extensive-stage) and metastasis status. Patients were excluded if aged younger than 60 years, or if diagnostic or treatment records, pathological or imaging confirmation, sex information or bone metastasis data were missing, or if the date of bone metastasis diagnosis was not consistent with the initial diagnosis of the primary disease (Figure 1). Applying the same age and histological criteria to the SEER cohort, we excluded 9 patients with missing race, 28 with histological grade I or II, 98 with T0 disease and 428 with missing bone, brain or liver metastasis fields, leaving 19,561 patients for analysis. The two hospitals ultimately contributed 1,387 patients, randomly divided 7:3 into a training set (n = 971) and an internal validation set (n = 416). The outcome was bone metastasis detected at initial diagnosis (yes or no): in the hospital cohorts this was determined from medical records and imaging reports, and in the SEER cohort the recorded bone metastasis field at diagnosis was used. Some

variables were missing in the retrospective data: in the two-hospital cohort missingness was greatest for tumour size (149 patients), primary site (68) and histological grade (29), while no other variable had more than 8 missing values; in the SEER cohort missingness was greatest for tumour size (30.25%), primary site (18.77%), T stage (17.15%) and the bone metastasis field (5.18%). All missing values were imputed by multiple imputation using chained equations with the mice package in R (m = 5, random seed 123), and the second completed dataset was carried forward for analysis.

Candidate variables

Sixteen candidate predictors were considered, all routine clinical and staging variables available at initial diagnosis: age, sex, primary site, laterality, T stage, N stage, clinical stage, SCLC stage (limited-stage or extensive-stage), histological grade, residence, liver metastasis, surgery, chemotherapy, lung metastasis, brain metastasis and tumour size. Lung and brain metastasis reflect metastatic status at initial diagnosis and were described in the baseline characteristics only and not entered into feature selection or modelling: like bone metastasis they are metastatic information recorded at initial diagnosis, so including them directly would create structural overlap with the outcome, and the recording conventions for these two fields are not entirely consistent between domestic and international data sources. Fourteen candidate variables therefore entered Boruta feature selection.

Feature selection

The Boruta algorithm was used to select variables [16]. For every real feature the algorithm generates randomly shuffled shadow features, compares real and shadow features through random forest feature importance (Z-scores), and after 100 iterations classifies as confirmed those variables whose importance is consistently higher than the maximum shadow feature importance. Only confirmed features were used as inputs for the subsequent models.

Model development

Six models were fitted on the training set: logistic regression (LR), decision tree (DT), random forest (RF), XGBoost, LightGBM and artificial neural network (ANN). LR and ANN standardised the input after dummy encoding, whereas the tree-based algorithms used the original categorical inputs directly. Hyperparameters were determined by 5-fold cross-validation, and the optimal parameter sets (Additional file 1: Table S1) were used uniformly for evaluation. LightGBM was implemented within the gradient boosting decision tree framework [17].

Model evaluation and clinical utility

Discrimination was measured by the area under the receiver operating characteristic curve (AUC) with 95% confidence intervals estimated by bootstrapping, and accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score and Cohen kappa were calculated at a fixed probability cut-off of 0.5. Calibration was assessed with calibration curves, together with the calibration intercept, calibration slope and Brier score [18]. Clinical utility was evaluated by decision curve analysis (DCA), comparing the net benefit of the model with the treat-all and treat-none reference strategies [19]. Differences in AUC between models were compared by paired bootstrap resampling (1,000 iterations); the multi-metric performance of the models in the internal validation set, the calibration parameters of LightGBM and the pairwise AUC comparisons are presented in Additional files 6, 7 and 8: Tables S6, S7 and S8, respectively. The representative model was chosen on the basis of discrimination stability across the four cohorts, calibration, and the feasibility of interpretability analysis, rather than the highest AUC in any single cohort.

Interpretability analysis

SHAP was applied to the representative model [20]. At the global level we plotted the feature importance bar chart ranked by mean absolute SHAP value and the beeswarm plot to show the magnitude and direction of each variable contribution; SHAP dependence plots were drawn for each input feature to examine its relationship with predicted risk; and at the individual level a representative case was selected for a waterfall plot that reconstructs the contribution of each item to a single prediction.

Statistical analysis and ethics

Statistical analysis was performed in R and Python 3.10. Missing values were imputed by multiple imputation using chained equations with the mice package in R (m = 5, random seed 123), and features were selected with the Boruta algorithm (Boruta package in R, 100 iterations). All candidate variables were categorical and are presented as counts (%), with groups compared using the chi-square test or Fisher exact test. Model fitting and evaluation (discrimination, calibration and decision curve analysis) and SHAP interpretation were performed in Python with scikit-learn, statsmodels, lightgbm, xgboost and shap; 95% CIs of the AUC were estimated by bootstrap resampling, and differences in AUC between models were compared by paired bootstrap resampling (1000 replicates). All tests were two-sided, and $P < 0.05$ was considered statistically significant. The study protocols of both hospitals were approved by the ethics committees of the respective institutions, and informed consent was waived because the study used retrospective anonymised data; SEER is a publicly available, de-identified database and therefore did not require ethical review.

Results

Study population and baseline characteristics

After screening according to the flow diagram in Figure 1, the two-hospital development cohort comprised 1,387 elderly patients with SCLC, of whom 313 (22.57%) had bone

metastasis. Bone metastasis was present in 234 of 971 patients (24.10%) in the training set and 79 of 416 (18.99%) in the internal validation set; the difference between the two groups was statistically significant (chi-square = 4.35, P = 0.037) and was attributed to proportional fluctuation arising from random splitting. External validation cohort 1 comprised 128 patients (29 with bone metastasis, 22.66%) and external validation cohort 2 comprised 19,561 patients (4,674 with bone metastasis, 23.89%); these rates were close to that of the development cohort, so the four cohorts shared a similar background prevalence of the outcome. Table 1 compares the distribution of age, sex, primary site, laterality, T stage, N stage, clinical stage, SCLC stage, histological grade, residence, lung metastasis, liver metastasis, brain metastasis, surgery, chemotherapy and tumour size between the training and internal validation sets, and no variable differed significantly between the two groups (all P > 0.05). In the overall population, 79.52% were male, 58.04% had extensive-stage disease, clinical stage III (46.79%) and IV (47.01%) predominated, 58.98% were N2, and 92.00% received chemotherapy; the most common tumour size category was less than 3 cm (43.84%), followed by 3-5 cm (36.27%). Baseline comparisons by bone metastasis status within each cohort are given in Additional files 2-4: Tables S2-S4, and the overall baseline characteristics of the four cohorts in Additional file 5: Table S5.

Feature selection

The results of applying Boruta to the 14 candidate variables (the 16 candidates minus lung and brain metastasis, which were described at baseline only) are shown in Figure 2. The box plot (Figure 2A) shows that the importance Z-scores of four variables - N stage, SCLC stage, liver metastasis and clinical stage - exceeded the threshold set by the maximum shadow feature importance and were classified as confirmed features, with liver metastasis having the highest overall importance; the remaining ten variables (age, sex, primary site, laterality, T stage, histological grade, residence, surgery, chemotherapy and tumour size) had importance

close to or below that of the shadow features and were classified as rejected and removed from modelling. The importance trajectory plot (Figure 2B) shows that the four confirmed features maintained high and stable importance throughout the 100 iterations, whereas the rejected variables fluctuated around zero and were negative in some iterations, indicating that the selection was robust. The four confirmed features - N stage, SCLC stage (limited-stage or extensive-stage), liver metastasis and clinical stage - were therefore used as inputs to the machine learning models.

Predictive performance of the models

Using the four confirmed features as inputs, six models were built on the training set and their discrimination was evaluated in the internal validation set and the two external cohorts. The ROC curves are shown in Figure 3; the multi-metric performance of the models in the internal validation set is given in Additional file 6: Table S6, and the complete multi-metric results for the four cohorts in Additional file 9: Table S9. In the training set the AUC of the models ranged from 0.791 to 0.820, with random forest highest (0.820, 95% CI 0.791-0.845), followed by XGBoost (0.814), decision tree (0.813) and LightGBM (0.810), and the 95% confidence intervals of the models overlapped. In the internal validation set the AUC ranged from 0.780 to 0.802, with ANN highest (0.802, 95% CI 0.750-0.851) and LightGBM second (0.795, 95% CI 0.739-0.846). In external validation cohort 1 the AUC of the models other than LR ranged from 0.792 to 0.815, with XGBoost highest (0.815, 95% CI 0.734-0.885), random forest close behind (0.814) and LR lowest (0.762). In external validation cohort 2 (SEER) LR had the highest AUC (0.754, 95% CI 0.748-0.761), the tree ensemble models ranged from 0.705 to 0.723, and RF (0.705) and ANN (0.702) were lowest. Overall, no single algorithm was consistently superior across the four cohorts, and the models that led on AUC in the training set did not necessarily retain that advantage externally. At a probability cut-off of 0.5 the models displayed two performance patterns. The first was a high-specificity pattern,

seen for LR and LightGBM in the training and internal validation sets, for example LightGBM in the training set with sensitivity 32.9%, specificity 91.6% and NPV 81.1%, which suits confirmation of positive risk. The second was a high-sensitivity pattern, seen for most models in the external cohorts: in external validation cohort 1 every model except LR reached a sensitivity of 93.1% with NPV 96.7-96.9%, and in SEER, DT, RF, LightGBM and ANN reached a sensitivity of 100% with NPV close to 100% but specificity of only 38.1-38.9%. These patterns indicate that, by adjusting the probability cut-off, the same model can serve either to confirm high risk or to rule out low-risk patients. Taking into account the fluctuation in AUC across the four cohorts together with calibration and interpretability requirements, LightGBM was chosen as the representative model: its AUCs in the training set, the internal validation set, external validation cohort 1 and cohort 2 were 0.810, 0.795, 0.807 and 0.718, respectively, the smallest fluctuation across cohorts and the best overall balance.

Model calibration and clinical utility

Calibration curves were plotted for LightGBM in the four cohorts (Figure 4). The calibration curve in the training set coincided almost exactly with the ideal diagonal (calibration intercept 0.026, slope 1.040); calibration in the internal validation set was good (intercept -0.411, slope 0.879) with mild systematic overestimation. Brier scores were 0.140 (95% CI 0.129-0.152) in the training set, 0.130 (95% CI 0.114-0.148) in the internal validation set, 0.182 (95% CI 0.147-0.219) in external validation cohort 1 and 0.234 (95% CI 0.231-0.237) in external validation cohort 2. The calibration curves of the two external cohorts deviated from the ideal line (mildly in cohort 1 and markedly in cohort 2), indicating drift in the probability scale when the model is applied across populations; because the intercept and slope were not annotated on the two external calibration plots, their calibration is described here only through the Brier score and the shape of the curve. The calibration intercept, slope and Brier score of LightGBM in the training and internal validation sets are given in

Additional file 7: Table S7. Decision curve analysis (Figure 5) showed that across the clinically relevant range of threshold probabilities the net benefit of LightGBM in all four cohorts exceeded that of both reference strategies, treat all and treat none; in the external cohorts, and particularly in SEER, the range of net benefit narrowed, consistent with the decline in discrimination and calibration as the population changes.

SHAP interpretability analysis

The global feature importance bar chart (Figure 6A), ranked by mean absolute SHAP value, placed clinical stage first (0.14), followed by SCLC stage (0.04) and liver metastasis (0.04) with comparable magnitude, while N stage contributed relatively little (0.02). The beeswarm plot (Figure 6B) shows the direction of each feature effect: the higher the clinical stage, the larger the positive SHAP value, that is, higher stage increases the predicted risk of bone metastasis; extensive-stage disease corresponded to a positive contribution and limited-stage disease to a negative one; patients with liver metastasis had positive SHAP values and those without had negative values; and higher N stage showed a positive trend, though of smaller magnitude than the first three. The SHAP dependence plots (Figure 7) further showed that the four features were positive overall: SCLC stage and liver metastasis were monotonically positive, that is, SHAP values rose steadily from limited-stage to extensive-stage disease and from absence to presence of liver metastasis, whereas clinical stage and N stage were positive overall but fluctuated somewhat at intermediate values, indicating that their relationship with predicted risk is not strictly linear. At the individual level, for a patient in the training set with a predicted probability of 0.37 (above the cohort baseline of 0.24), the waterfall plot (Figure 8) showed that the main positive contributions came from higher clinical stage and extensive-stage status, while N stage contributed little, suggesting that this individual risk was driven mainly by tumour burden and extensive-stage features.

Discussion

In this study we developed and compared six prediction models in 1,387 elderly patients with SCLC and then performed external validation in an independent Chinese hospital cohort and a cross-population SEER cohort. The AUCs of the six models in the training set, the internal validation set, external validation cohort 1 and cohort 2 ranged from 0.79 to 0.82, 0.78 to 0.80, 0.76 to 0.82 and 0.70 to 0.75, respectively, showing a pattern of slightly higher performance internally and progressive decline externally, and the best-performing algorithm differed between cohorts. LightGBM was selected as the representative model because it fluctuated least across cohorts, and its AUC in external validation cohort 2 (0.718) was in line with that of the other models in the same cohort, suggesting that the overall fall in external performance is mainly attributable to population differences rather than to an inappropriate choice of algorithm.

The final model retained only four features: clinical stage, SCLC stage, N stage and liver metastasis. The TNM clinical stage of SCLC has prognostic discriminatory value in real-world data [21], and the extensive-stage versus limited-stage dichotomy is the backbone of clinical treatment decisions; concurrent liver metastasis at initial diagnosis indicates that the tumour already has the capacity for haematogenous multi-organ dissemination, which frequently coexists with bone metastasis as synchronous metastasis. Mechanistic studies have repeatedly emphasised the relationship between early dissemination in SCLC and the plasticity of the neuroendocrine phenotype [22,23], which corroborates the dominance of staging-related variables in the model. The rejection by Boruta of non-tumour-burden variables such as age, sex and residence also suggests that, at the point of initial diagnosis, information reflecting the extent of tumour invasion characterises the risk of bone metastasis better than demographic information does.

Compared with previous studies predicting bone metastasis in SCLC, the differences in the present work lie mainly in the data and the validation strategy. Nomograms and machine learning models for bone metastasis based on the SEER database have mostly relied on internal resampling for validation [12,13]. In this study we trained on data from two hospitals and, in addition to SEER, established a separate Chinese hospital cohort, allowing the transferability of the model to be examined both across institutions and across populations. In addition, prognostic or predictive tools specific to brain metastasis and to liver metastasis in SCLC have already been reported [24,25], which suggests that this field tends to model metastasis sites and outcomes separately; the outcome definitions and populations of studies on bone, brain and liver metastasis are not the same, so their performance figures cannot be compared directly.

At the level of algorithm choice, our results echo the conclusions of previous systematic reviews: for most clinical prediction problems, complex machine learning is not necessarily superior to a well-specified logistic regression [26]. In this study LR achieved the highest AUC in the SEER cohort (0.754), whereas RF was best in the training set yet worst in the SEER cohort, making the trade-off between overfitting and generalisation plainly visible. Selecting and reporting a model solely on the basis of development-set AUC is likely to overestimate its cross-population performance; presenting the distribution of performance across cohorts and disclosing the calibration of probabilities is closer to the practical needs of clinical decision-making than pursuing the best value of a single metric [27].

Discrimination and calibration should be treated on an equal footing [18]. LightGBM was well calibrated in the training and internal validation cohorts (Brier scores 0.140 and 0.130) but deviated from the ideal line in the two external cohorts (Brier scores 0.182 and 0.234). This is precisely the point of external validation: a model can borrow the feature-outcome relationship learned from the training data, but it cannot automatically inherit the outcome

baseline of the training population. The prevalence of bone metastasis in the SEER cohort (23.89%) was comparable to that in the development cohort (22.57%), indicating that the drift does not arise from a difference in prevalence but more probably from differences in variable definitions, the level of imaging detection and population composition [27]. Before a model is transferred to a new institution, therefore, local samples should be used to recalibrate the intercept and even the slope.

Interpretability is a prerequisite for a model to enter clinical practice. SHAP analysis showed that clinical stage contributed most to prediction, that extensive-stage disease, concurrent liver metastasis and higher N stage all pushed predictions towards the high-risk side, and that the composition of contributions for a single prediction can be reconstructed item by item, which makes it easier for clinicians to understand the output and explain it to patients. The mathematical basis of SHAP has been examined for its applicability across different prediction tasks [28]. In practical terms, the most direct use of risk stratification is to triage imaging resources: low-risk patients can have whole-body bone scintigraphy and PET/CT reduced, while high-risk patients should promptly undergo complete staging imaging and be assessed for bone-modifying agents and palliative radiotherapy, thereby reducing the risk of pathological fracture and spinal cord compression.

Prediction should also be understood against the background of evolving treatment. Immunotherapy combined with chemotherapy has become the first-line choice for extensive-stage SCLC, and real-world data suggest that elderly patients can benefit as well, with poor performance status or a high inflammatory burden not constituting an absolute contraindication [29-31]. The widening range of treatment options gives more weight to the judgement of whether additional staging for bone metastasis is worthwhile: once bone metastasis is present it affects not only systemic treatment but also the management of local symptoms and the arrangement of supportive care, and the growing discussion of surgical

intervention in early-stage SCLC likewise increases the value of accurate staging [32]. Future work could attempt to integrate CT radiomic features with clinical variables [33] and to compare the robustness of different feature selection algorithms in larger, multicentre datasets [34].

Limitations

This study has several limitations. First, the retrospective design relies on medical records and database coding, and the determination of bone metastasis could not be verified by central imaging review; the proportion of screened patients excluded from the two hospital cohorts was relatively high, which may introduce selection bias. Second, the outcome was restricted to bone metastasis detected at initial diagnosis and did not cover metastases arising after diagnosis, which are equally important for treatment during follow-up. Third, random splitting produced a difference in the prevalence of bone metastasis between the training and internal validation sets (24.10% versus 18.99%, P = 0.037); this reflects random fluctuation in a sample with few events, but it indicates that outcome-stratified splitting should be considered when the sample is limited. Fourth, the candidate variables were confined to routine clinical and staging information and did not include smoking dose, laboratory indices or quantitative imaging features; lung and brain metastasis were not entered into the model as co-metastatic information, so the model does not fully capture patterns of synchronous multi-site dissemination. Fifth, the cohort from the First Affiliated Hospital was small (128 patients), and the SEER cohort differs from the Chinese cohorts in population composition and data coding; although differences in AUC between models were assessed by paired bootstrap testing (Additional file 8: Table S8), the limited size of the cohorts means that differences between external cohorts may not have been detected, and calibration in the external cohorts was described only through the Brier score and the shape of the curve. These limitations suggest that the conclusions of this study are better framed as describing an

interpretable risk stratification tool that requires recalibration, and that its ultimate clinical value still depends on prospective, multicentre external validation.

Conclusions

This study used bone metastasis at initial diagnosis as the outcome and developed and compared six prediction models in elderly patients with SCLC on the basis of four clinical features available at initial diagnosis. LightGBM showed stable discrimination, clear calibration behaviour and interpretable output in the training set, the internal validation set and the two external cohorts, and can serve as an auxiliary tool for the early stratification of bone metastasis risk and for the allocation of staging imaging resources; probability recalibration using local populations should be performed before cross-institutional application, and its clinical value awaits confirmation in prospective multicentre data.

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committees of the Chinese PLA General Hospital (approval number 2026KY164-KS001), Peking University Cancer Hospital (approval number 2025KT190) and the First Affiliated Hospital of Hebei North University (approval number HBNV20250322070). Because the study used retrospective, anonymised data, the requirement for written informed consent was waived. The SEER database contains publicly available, de-identified data and therefore did not require ethical approval or informed consent.

Consent for publication

Not applicable.

Availability of data and materials

The SEER data analysed in this study are publicly available from the Surveillance, Epidemiology, and End Results programme of the National Cancer Institute. The data from the two development hospitals and from the external Chinese validation hospital are not publicly available because of institutional ethical restrictions, but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by the National Natural Science Foundation of China Youth Science Fund Project (No.12404501) and the Hebei Provincial Natural Science Foundation Project (H2025405013).

Authors' contributions

D.S., Y.Z. and Z.M. contributed equally to this manuscript. D.S. and Y.Z. conceived of the study, participated in its design and drafted the manuscript. D.S. and Z.M. participated in the methodology and revised the manuscript critically. Y.Z. and Z.M. provided the resources. Y.Z., Z.M., J.Y. and H.C. performed the investigation. J.Y. and H.C. performed the formal analysis. W.H., B.W. and W.L. supervised the study and acquired the funding. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Additional files

Additional file 1: Table S1. Parameter settings and optimization strategies for the six machine learning models. File format: DOCX (.docx).

Additional file 2: Table S2. Comparison of clinical characteristics between the bone metastasis group and the non-bone metastasis group in the two-hospital development cohort. File format: DOCX (.docx).

Additional file 3: Table S3. Comparison of clinical characteristics between the bone metastasis group and the non-bone metastasis group in the SEER cohort. File format: DOCX (.docx).

Additional file 4: Table S4. Comparison of clinical characteristics between the bone metastasis group and the non-bone metastasis group in the first affiliated hospital external validation cohort. File format: DOCX (.docx).

Additional file 5: Table S5. Baseline demographic and clinicopathological characteristics across the training, internal validation and external validation cohorts. File format: DOCX (.docx).

Additional file 6: Table S6. Comparison of prediction performance indicators of each machine learning model in the internal validation set (AUC, accuracy, sensitivity, specificity, F1). File format: DOCX (.docx).

Additional file 7: Table S7. Calibration performance of the LightGBM model in the training and internal validation cohorts. File format: DOCX (.docx).

Additional file 8: Table S8. Pairwise bootstrap comparison of AUC performance between LightGBM and the other machine learning models in the internal validation cohort. File format: DOCX (.docx).

Additional file 9: Table S9. Detailed prediction performance of the six machine learning models across the training set, the internal validation set and the two external validation cohorts. File format: DOCX (.docx).

Additional file 10: Supplementary methods. Detailed technical specifications of the data sources, variable coding, missing-data handling, feature selection, model hyperparameters and interpretability analyses. File format: DOCX (.docx).

References

1. Rudin CM, Brambilla E, Faivre-Finn C, Sage J. Small-cell lung cancer. Nat Rev Dis Primers. 2021;7(1):3.

2. Turner MC, Andersen ZJ, Baccarelli A, Diver WR, Gapstur SM, Pope CA, et al. Outdoor air pollution and cancer: an overview of the current evidence and public health recommendations. CA Cancer J Clin. 2020;70(6):460-479.

3. Aarts MJ, Aerts JG, van den Borne BE, Biesma B, Lemmens VE, Kloover JS. Comorbidity in patients with small-cell lung cancer: trends and prognostic impact. Clin Lung Cancer. 2015;16(4):282-291.

4. Zou J, Guo S, Xiong MT, Xu Y, Shao J, Tong Z, et al. Ageing as key factor for distant metastasis patterns and prognosis in patients with extensive-stage small cell lung cancer. J Cancer. 2021;12(6):1575-1582.

5. Oronsky B, Abrouk N, Caroen S, Lybeck M, Guo X, Wang X, et al. A 2022 update on extensive stage small-cell lung cancer (SCLC). J Cancer. 2022;13(9):2945-2953.

6. Riihimaki M, Hemminki A, Fallah M, Thomsen H, Sundquist K, Sundquist J, et al. Metastatic sites and survival in lung cancer. Lung Cancer. 2014;86(1):78-84.

7. Cai H, Wang H, Li Z, Lin J, Yu J. The prognostic analysis of different metastatic patterns in extensive-stage small-cell lung cancer patients: a large population-based study. Future Oncol. 2018;14(14):1397-1407.

8. Megyesfalvi Z, Tallosy B, Pipek O, Fillinger J, Lang C, Klikovits T, et al. The landscape of small cell lung cancer metastases: organ specificity and timing. Thorac Cancer. 2021;12(6):914-923.

9. Wu Y, Zhang J, Zhou W, Yuan Z, Wang H. Prognostic factors in extensive-stage small cell lung cancer patients with organ-specific metastasis: unveiling commonalities and disparities. J Cancer Res Clin Oncol. 2024;150(2):74.

10. Ganti AKP, Loo BW, Bassetti M, Blakely C, Chiang A, D'Amico TA, et al. Small cell lung cancer, version 2.2022, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2021;19(12):1441-1464.

11. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. Am J Epidemiol. 2019;188(12):2222-2239.

12. Liu C, Yi J, Jia J. Diagnostic and prognostic nomograms for bone metastasis in small cell lung cancer. J Int Med Res. 2021;49(10):3000605211050735.

13. Qie S, Zhang X, Luan J, Song Z, Li J, Wang J. Model development and validation for predicting small-cell lung cancer bone metastasis utilizing diverse machine learning algorithms based on the SEER database. Medicine (Baltimore). 2025;104(12):e41987.

14. Chen H, Dai W, Yang J, Dang X, Jiang L. Interpretable machine learning-based prediction of liver metastasis risk in elderly patients with small cell lung cancer: a study based on the SEER database and external validation in a Chinese cohort. Int J Med Inform. 2026;209:106274.

15. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378.

16. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw. 2010;36(11):1-13.

17. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems 30 (NeurIPS 2017). Long Beach, CA; 2017. p. 3146-3154.

18. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230.

19. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-574.

20. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2(1):56-67.

21. Arriola E, Trigo JM, Sánchez-Gastaldo A, Navarro A, Perez C, Crama L, et al. Prognostic value of clinical staging according to TNM in patients with SCLC: a real-world surveillance, epidemiology and end-results database analysis. JTO Clin Res Rep. 2022;3(1):100266.

22. Ko J, Winslow MM, Sage J. Mechanisms of small cell lung cancer metastasis. EMBO Mol Med. 2021;13(1):e13122.

23. Jin Y, Chen Y, Qin Z, Hu L, Guo C, Ji H. Understanding SCLC heterogeneity and plasticity in cancer metastasis and chemotherapy resistance. Acta Biochim Biophys Sin (Shanghai). 2023;55(6):948-955.

24. Rong YT, Zhu YC, Wu Y. A novel nomogram predicting cancer-specific survival in small cell lung cancer patients with brain metastasis. Transl Cancer Res. 2022;11(12):4289-4302.

25. Li R, Li H, Tang C, Cui H, Zhong R, Liu Y, et al. Metastatic sites, prognosis, and a new nomogram for predicting overall survival among small cell lung cancer patients with liver metastasis: a retrospective study based on SEER. J Thorac Dis. 2025;17(1):344-356.

26. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12-22.

27. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, how, why, when and where? Clin Kidney J. 2021;14(1):49-58.
28. Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using Shapley values: application to compound potency and multi-target activity predictions. J Comput Aided Mol Des. 2020;34(10):1013-1026.
29. Zhao K, Lu S, Niu J, Zhu H, Tian Y, Yu J. Real-world data on immunotherapy combined with chemotherapy in elderly patients with extensive-stage small cell lung cancer. BMC Cancer. 2025;25(1):467.
30. Li X, Tong L, Wang S, Yu J, Lu B, Wang Q, et al. Integration of clinical and blood parameters in risk prognostication for patients receiving immunochemotherapy for extensive stage small cell lung cancer: real-world data from two centers. BMC Med. 2024;22(1):381.
31. Inomata M, Takata N, Seto Z, Murayama N, Tokui K, Okazawa S, et al. Effectiveness of chemoimmunotherapy in small-cell lung cancer patients with a poor performance status or higher neutrophil/lymphocyte ratio. In Vivo. 2025;39(1):467-472.
32. Sezer HF. Current surgical perspective on the prognosis of small-cell lung cancer. Diagnostics (Basel). 2025;15(21):2704.
33. Wang F, Chen W, Chen F, Lu J, Xu Y, Fang M, et al. Risk stratification and overall survival prediction in extensive stage small cell lung cancer after chemotherapy with immunotherapy based on CT radiomics. Sci Rep. 2024;14(1):22659.
34. Mohtasham F, Pourhoseingholi MA, Hashemi Nazari SS, Kavousi K, Zali MR. Comparative analysis of feature selection techniques for COVID-19 dataset. Sci Rep. 2024;14(1):18627.

Figure legends

Figure 1. Flowchart of patient selection and model development. Elderly patients (aged 60 years or older) with small cell lung cancer diagnosed between 2010 and 2024 were

retrospectively enrolled from the Chinese PLA General Hospital and Peking University Cancer Hospital (N = 8,596); after applying the exclusion criteria, 1,387 patients were included and randomly split 7:3 into a training set (n = 971) and an internal validation set (n = 416). External validation cohort 1 comprised patients diagnosed at the First Affiliated Hospital of Hebei North University between 2010 and 2024 (N = 1,506 screened; 128 included). External validation cohort 2 was derived from the SEER database (2010-2015; N = 20,124 screened); after excluding 9 patients with incomplete race information, 28 with histological grade I-II, 98 with T0 stage and 428 with incomplete bone, brain or liver metastasis information, 19,561 patients were included. Six machine learning models were developed and assessed with eight performance measures. SCLC, small cell lung cancer; SEER, Surveillance, Epidemiology, and End Results.

Figure 2. Boruta feature selection. (A) Box plot of variable importance for each candidate variable and for the shadow features: blue denotes the shadow features (shadowMin, shadowMean and shadowMax), green the confirmed features (N stage, SCLC stage, liver metastasis and clinical stage), yellow the tentative features and red the rejected features; the importance of a real feature must exceed the maximum shadow feature importance for that feature to be judged important. (B) Trajectory of variable importance over 100 iterations; the confirmed features remained consistently high throughout, whereas the rejected variables fluctuated around zero and were negative in some iterations, indicating that the selection was robust. SCLC, small cell lung cancer.

Figure 3. ROC curves of the six machine learning models in the four cohorts. (A) Training set; (B) internal validation set; (C) external validation cohort 1 (the First Affiliated Hospital of Hebei North University); (D) external validation cohort 2 (SEER database). The legend gives the AUC and 95% CI of each model. Random forest was highest in the training set (AUC 0.820) and ANN was highest in the internal validation set (0.802), whereas XGBoost and

585 random forest led in the two external cohorts; AUCs ranged from 0.702 to 0.820 across the
586 cohorts and no single model was consistently superior. ROC, receiver operating characteristic;
587 AUC, area under the receiver operating characteristic curve; CI, confidence interval; ANN,
588 artificial neural network; SEER, Surveillance, Epidemiology, and End Results.

589 Figure 4. Calibration curves of the LightGBM model in the four cohorts. The x-axis shows
590 the predicted probability and the y-axis the observed probability; the grey dashed line is the
591 line of ideal calibration and the blue points are the observed probabilities within each group
592 of predicted probabilities. Brier scores: (A) training set, 0.140 (95% CI 0.129-0.152); (B)
593 internal validation set, 0.130 (95% CI 0.114-0.148); (C) external validation cohort 1, 0.182
594 (95% CI 0.147-0.219); (D) external validation cohort 2, 0.234 (95% CI 0.231-0.237).
595 Calibration was good in the training and internal validation sets, whereas both external
596 cohorts showed appreciable drift in the probability scale. CI, confidence interval.

597 Figure 5. Decision curve analysis of the LightGBM model in the four cohorts. (A) Training
598 set; (B) internal validation set; (C) external validation cohort 1; (D) external validation cohort
599 2. The x-axis is the threshold probability and the y-axis the net benefit; the blue solid line is
600 the net benefit of the model, the red solid line the treat-all strategy and the green dashed line
601 the treat-none strategy. Across the clinically relevant range of threshold probabilities the net
602 benefit of the model exceeded that of both reference strategies in all four cohorts; the range of
603 net benefit narrowed in the external cohorts, particularly in SEER, consistent with the decline
604 in discrimination and calibration.

605 Figure 6. Global SHAP interpretation of the LightGBM model. (A) Feature importance bar
606 chart ranked by mean absolute SHAP value: clinical stage ranked first (0.14), followed by
607 SCLC stage (0.04) and liver metastasis (0.04) with comparable magnitude, while N stage
608 contributed relatively little (0.02). (B) Beeswarm plot, in which each point represents one
609 patient, the x-axis is the SHAP value and colour indicates whether the feature value is high or

low. The higher the clinical stage, the larger the positive SHAP value; extensive-stage disease, concurrent liver metastasis and higher N stage all contributed positively, whereas limited-stage disease and the absence of liver metastasis contributed negatively. SHAP, SHapley Additive exPlanations; SCLC, small cell lung cancer.

Figure 7. SHAP dependence plots for the four confirmed features. The x-axis is the feature value, the y-axis the SHAP value, colour indicates whether the feature value is high or low, and the orange dashed line marks SHAP = 0. SCLC stage and liver metastasis showed a monotonically positive trend, that is, SHAP values rose steadily from limited-stage to extensive-stage disease and from the absence to the presence of liver metastasis; clinical stage and N stage were positive overall but fluctuated somewhat at intermediate values, indicating that their relationship with the predicted risk is not strictly linear. SHAP, SHapley Additive exPlanations; SCLC, small cell lung cancer.

Figure 8. Individual SHAP waterfall plot for a representative patient in the training set (predicted probability 0.37, above the cohort baseline of 0.24). Red bars represent positive contributions that increase the predicted risk and blue bars negative contributions that decrease it. The main positive contributions came from higher clinical stage (+0.13) and extensive-stage status (+0.03), while N stage contributed little (-0.01), suggesting that this individual risk was driven mainly by tumour burden and extensive-stage features. SHAP, SHapley Additive exPlanations.

Table 1. Baseline characteristics of the training and internal validation cohorts.

Variables	**Total (n = 1387)**	**Internal validation (n = 416)**	**Training (n = 971)**	**Statistic**	***P*-value**
Bone metastasis				χ^2=4.35	0.037
No	1074 (77.43)	337 (81.01)	737		

Variables	Total (n = 1387)	Internal validation (n = 416)	Training (n = 971)	Statistic	*P-value*
			(75.90)		
Yes	313 (22.57)	79 (18.99)	234 (24.10)		
Primary site				χ^2=4.14	0.388
Upper lobe, lung	654 (47.15)	196 (47.12)	458 (47.17)		
Middle lobe, lung	136 (9.81)	39 (9.38)	97 (9.99)		
Lower lobe, lung	505 (36.41)	161 (38.70)	344 (35.43)		
Main bronchus	81 (5.84)	17 (4.09)	64 (6.59)		
Overlapping lesion of lung	11 (0.79)	3 (0.72)	8 (0.82)		
Laterality				-	0.699
Left	600 (43.26)	174 (41.83)	426 (43.87)		
Right	782 (56.38)	241 (57.93)	541 (55.72)		
Bilateral	5 (0.36)	1 (0.24)	4 (0.41)		
T stage				χ^2=0.23	0.973
T1	556 (40.09)	163 (39.18)	393 (40.47)		
T2	536 (38.64)	164 (39.42)	372		

Variables	Total (n = 1387)	Internal validation (n = 416)	Training (n = 971)	Statistic	*P-value*
			(38.31)		
T3	147 (10.60)	44 (10.58)	103 (10.61)		
T4	148 (10.67)	45 (10.82)	103 (10.61)		
Stage				χ^2=1.96	0.376
I/II	86 (6.20)	31 (7.45)	55 (5.66)		
III	649 (46.79)	197 (47.36)	452 (46.55)		
IV	652 (47.01)	188 (45.19)	464 (47.79)		
N stage				χ^2=1.74	0.628
N0	36 (2.60)	14 (3.37)	22 (2.27)		
N1	101 (7.28)	31 (7.45)	70 (7.21)		
N2	818 (58.98)	247 (59.38)	571 (58.81)		
N3	432 (31.15)	124 (29.81)	308 (31.72)		
Sex				χ^2=1.78	0.183
Female	284 (20.48)	76 (18.27)	208 (21.42)		
Male	1103 (79.52)	340 (81.73)	763		

Variables	Total (n = 1387)	Internal validation (n = 416)	Training (n = 971)	Statistic	*P-value*
			(78.58)		
Grade				χ^2=0.17	0.680
III	340 (24.51)	105 (25.24)	235 (24.20)		
IV	1047 (75.49)	311 (74.76)	736 (75.80)		
Residence				χ^2=1.31	0.252
Rural	589 (42.47)	167 (40.14)	422 (43.46)		
Urban	798 (57.53)	249 (59.86)	549 (56.54)		
SCLC stage				χ^2=1.00	0.316
LDS	582 (41.96)	183 (43.99)	399 (41.09)		
EDS	805 (58.04)	233 (56.01)	572 (58.91)		
Lung metastasis				χ^2=0.04	0.836
NO	1240 (89.40)	373 (89.66)	867 (89.29)		
YES	147 (10.60)	43 (10.34)	104 (10.71)		
Liver metastasis				χ^2=0.55	0.458

Variables	Total (n = 1387)	Internal validation (n = 416)	Training (n = 971)	Statistic	*P-value*
NO	1156 (83.35)	342 (82.21)	814 (83.83)		
YES	231 (16.65)	74 (17.79)	157 (16.17)		
Brain metastasis				χ^2=1.02	0.313
NO	1219 (87.89)	360 (86.54)	859 (88.47)		
YES	168 (12.11)	56 (13.46)	112 (11.53)		
Surgery				χ^2=0.03	0.870
NO	1342 (96.76)	403 (96.88)	939 (96.70)		
YES	45 (3.24)	13 (3.12)	32 (3.30)		
Chemotherapy				χ^2=1.03	0.309
NO	111 (8.00)	38 (9.13)	73 (7.52)		
YES	1276 (92.00)	378 (90.87)	898 (92.48)		
Tumor size				χ^2=1.65	0.800
0 ≤X<3	608 (43.84)	190 (45.67)	418 (43.05)		
3 ≤X<5	503 (36.27)	142 (34.13)	361 (37.18)		

Variables	Total (n = 1387)	Internal validation (n = 416)	Training (n = 971)	Statistic	*P-value*
5 ≤X<7	136 (9.81)	41 (9.86)	95 (9.78)		
7 ≤X<10	127 (9.16)	38 (9.13)	89 (9.17)		
X≥10	13 (0.94)	5 (1.20)	8 (0.82)		
Age				χ^2=1.35	0.930
60–64	494 (35.62)	152 (36.54)	342 (35.22)		
65–69	451 (32.52)	137 (32.93)	314 (32.34)		
70–74	277 (19.97)	79 (18.99)	198 (20.39)		
75–79	134 (9.66)	39 (9.38)	95 (9.78)		
80–84	24 (1.73)	6 (1.44)	18 (1.85)		
85+	7 (0.50)	3 (0.72)	4 (0.41)		

630

631 Data are presented as n (%). χ^2, chi-square test; -, Fisher exact test.

632

Two hospitals
2010-2024: patients diagnosed with small-cell lung cancer were retrospectively enrolled from Chinese PLA General Hospital and Peking University Cancer Hospital. (N = 8596)

Hospital 3
2010-2024: Patients diagnosed with small-cell lung cancer at The First Affiliated Hospital of Hebei North University (N = 1506)

Exclusion criteria
1) Age < 60 years
2) Missing diagnostic or treatment records
3) Missing pathological or imaging confirmation
4) Missing gender information
5) Bone metastasis data unavailable
6) The date of bone metastasis diagnosis was not consistent with the initial diagnosis of the primary disease

Final included elderly SCLC patients (N = 1387)

Final included elderly SCLC patients (N = 128)

Randomly split 7:3

Training Set (n = 971)

Internal Validation Set (n = 416)

External Validation Set 1

External Validation Set 2

Models
1)Logistic regression
2) LightGBM
3) Decision Tree
4) Random Forest
5) XGBoost
6) Artificial Neural Network

Evaluation measures
1) Accuracy
2) Sensitivity
3) Specificity
4) F1-score
5) PPV
6) AUC
7) Kappa
8) NPV

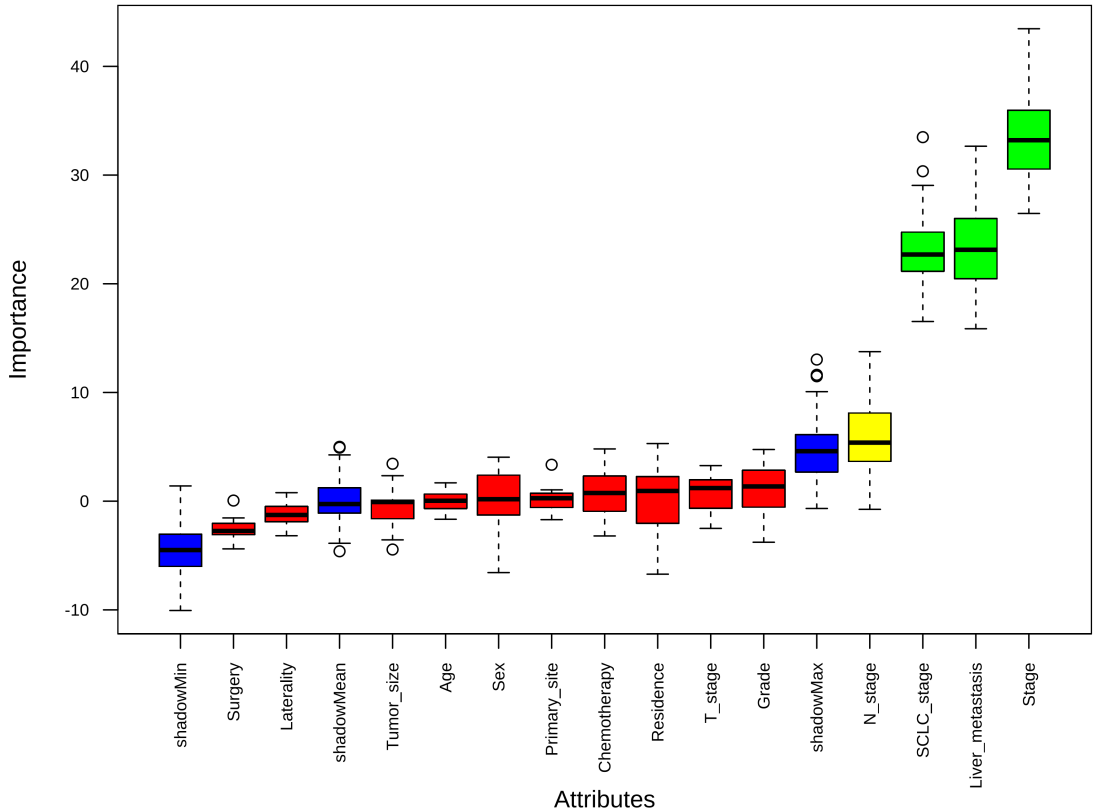
Elderly SCLC patients meeting inclusion criteria (N = 19561)

Incomplete information on race (n = 9)
Histological grade I–II (n = 28)
T0 stage without evidence of primary tumor (n = 98)
Incomplete information on bone metastasis (n = 4)
Incomplete information on brain metastasis (n = 1)
Incomplete information on liver metastasis (n = 423)

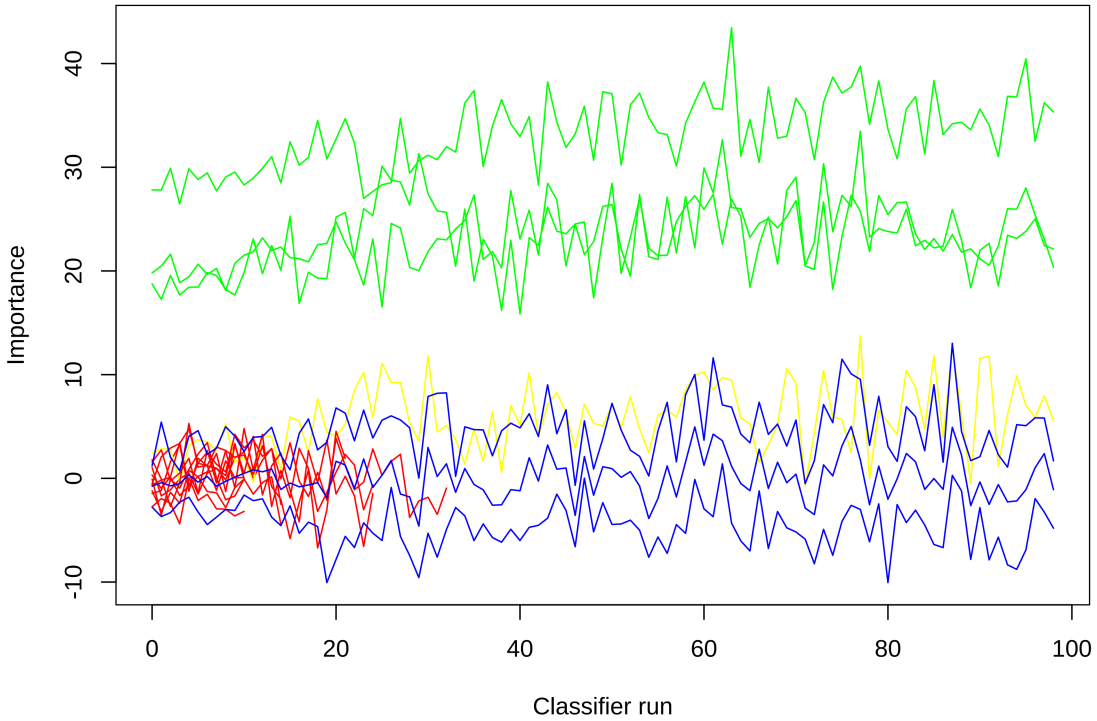
Patients aged ≥60 years diagnosed with small cell lung cancer in the SEER database (2010-2015) (N = 20124)

Fig. 1. The flowchart for patient selection and modeling.

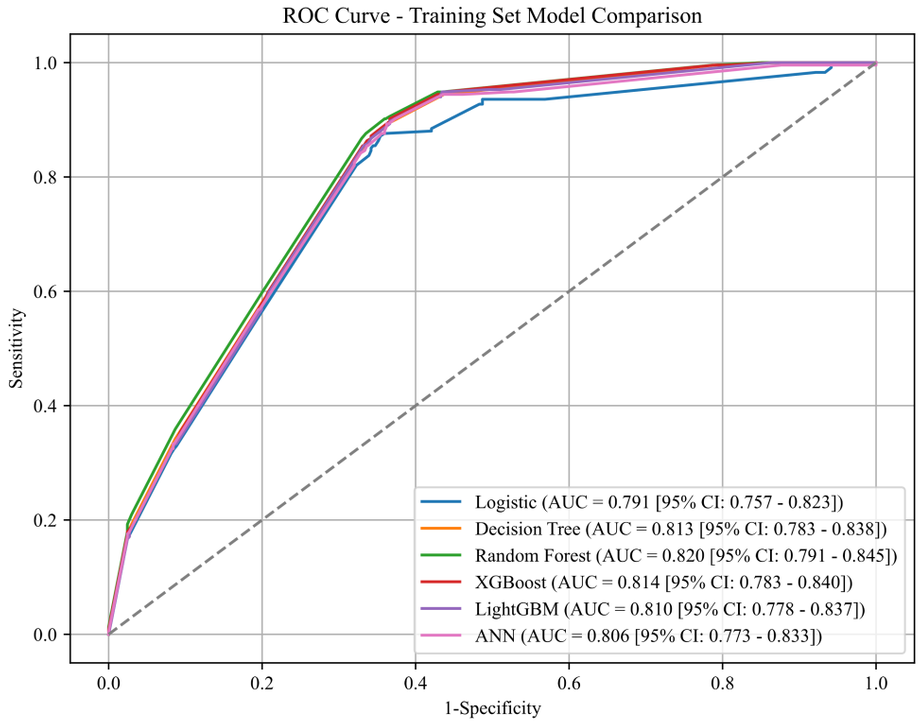
A


Importance
40
30
20
10
0
-10
shadowMin
Surgery
Laterality
shadowMean
Tumor_size
Age
Sex
Primary_site
Chemotherapy
Residence
T_stage
Grade
shadowMax
N_stage
SCLC_stage
Liver_metastasis
Stage
Attributes

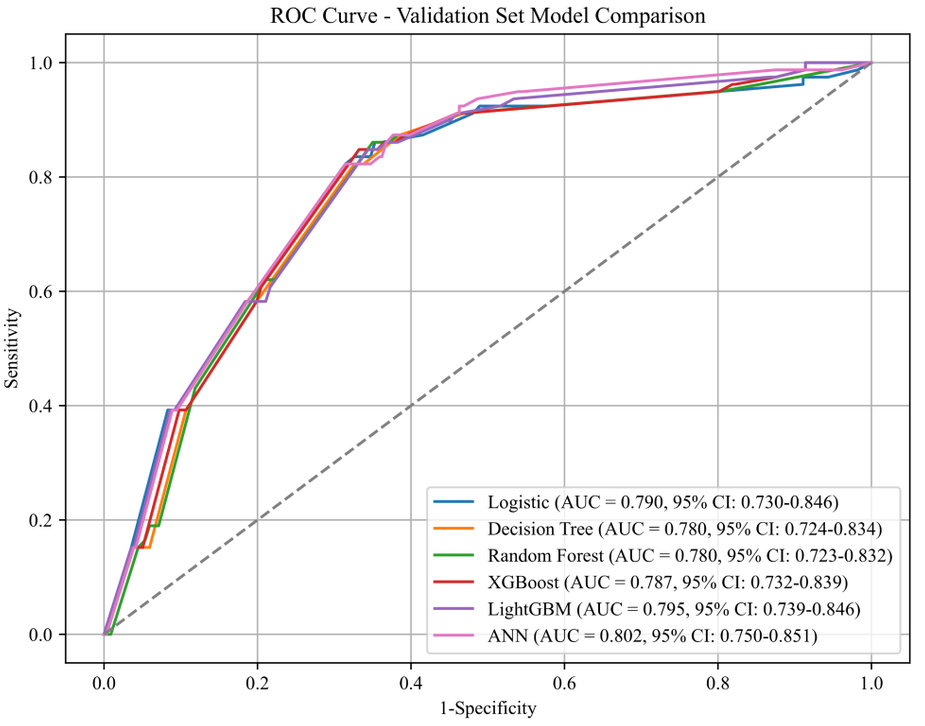
B


Importance
40
30
20
10
0
-10
0
20
40
60
80
100
Classifier run

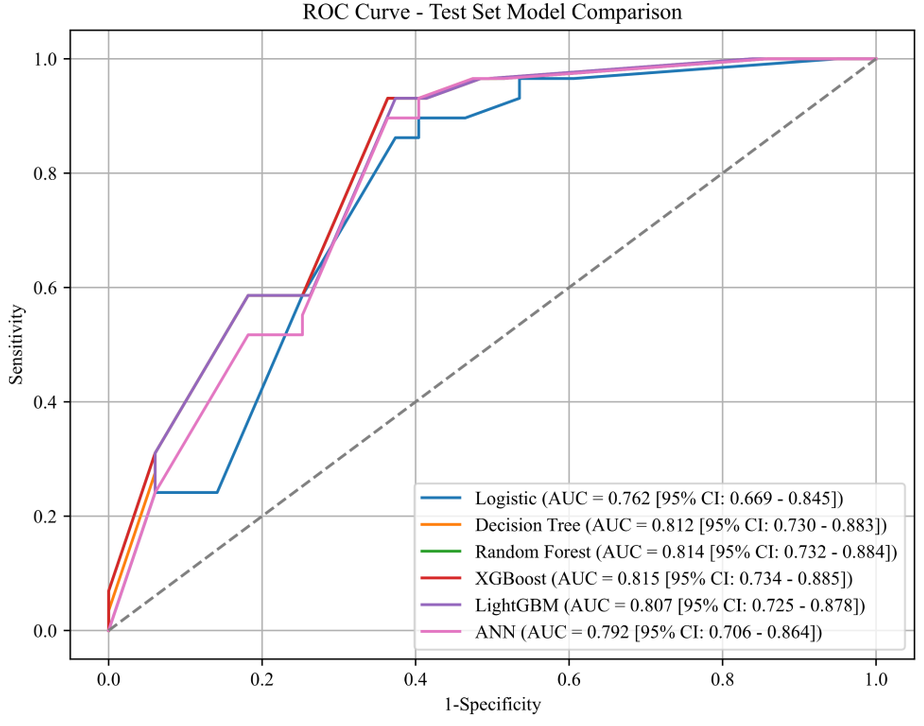
A Training set



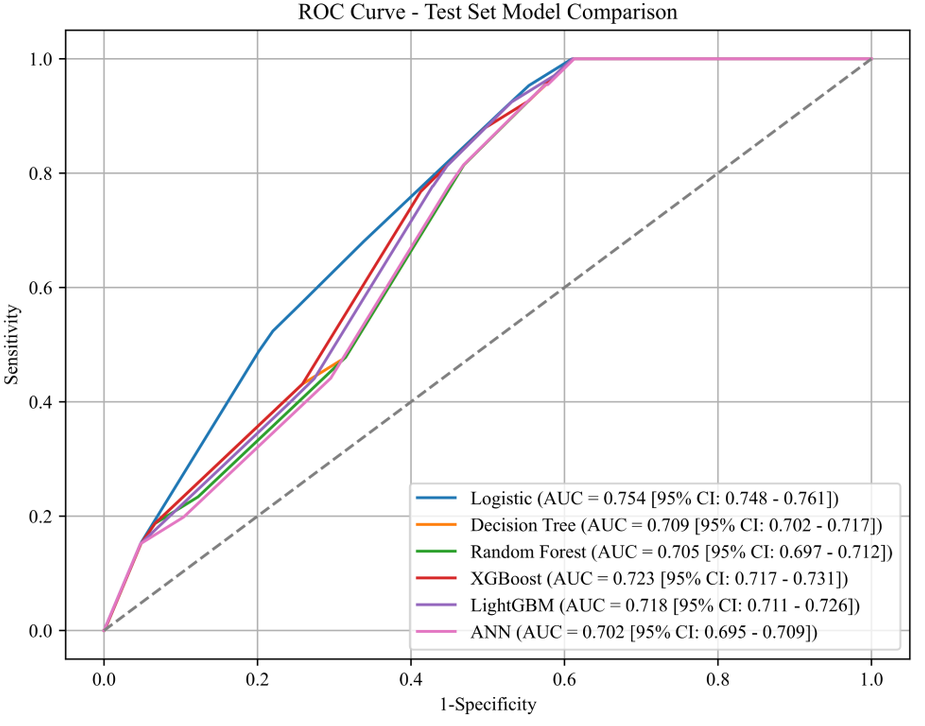
B Internal validation set



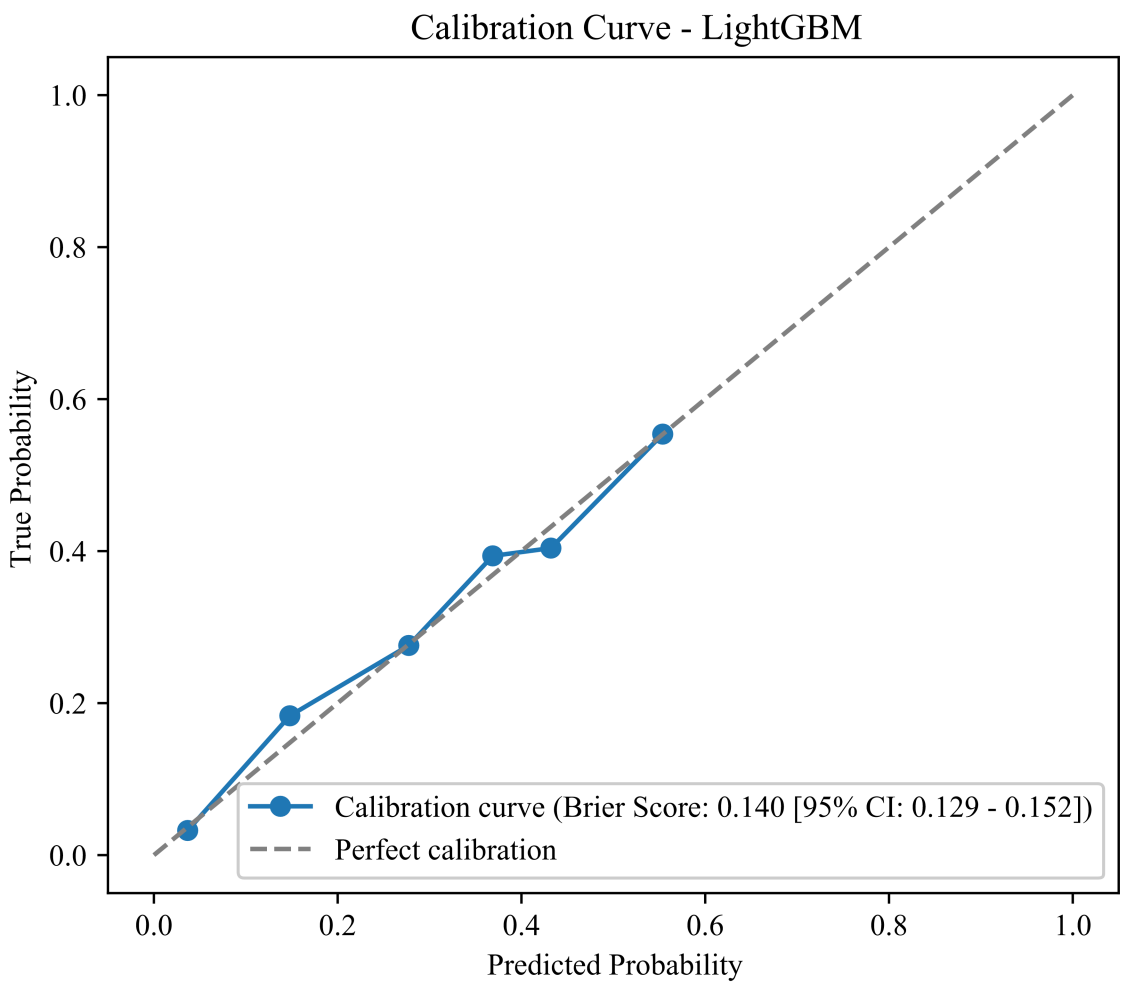
C External cohort 1 (Affiliated Hospital)



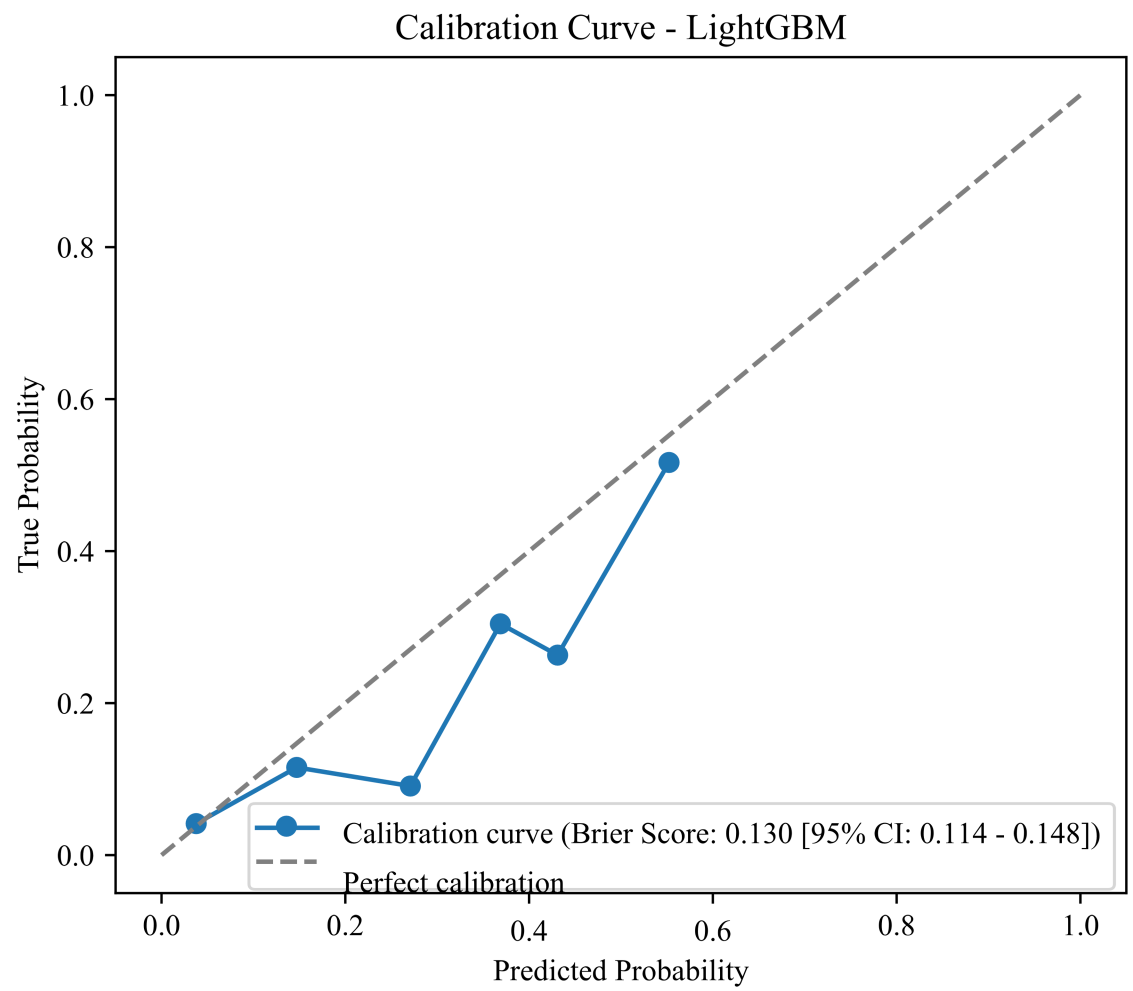
D External cohort 2 (SEER)



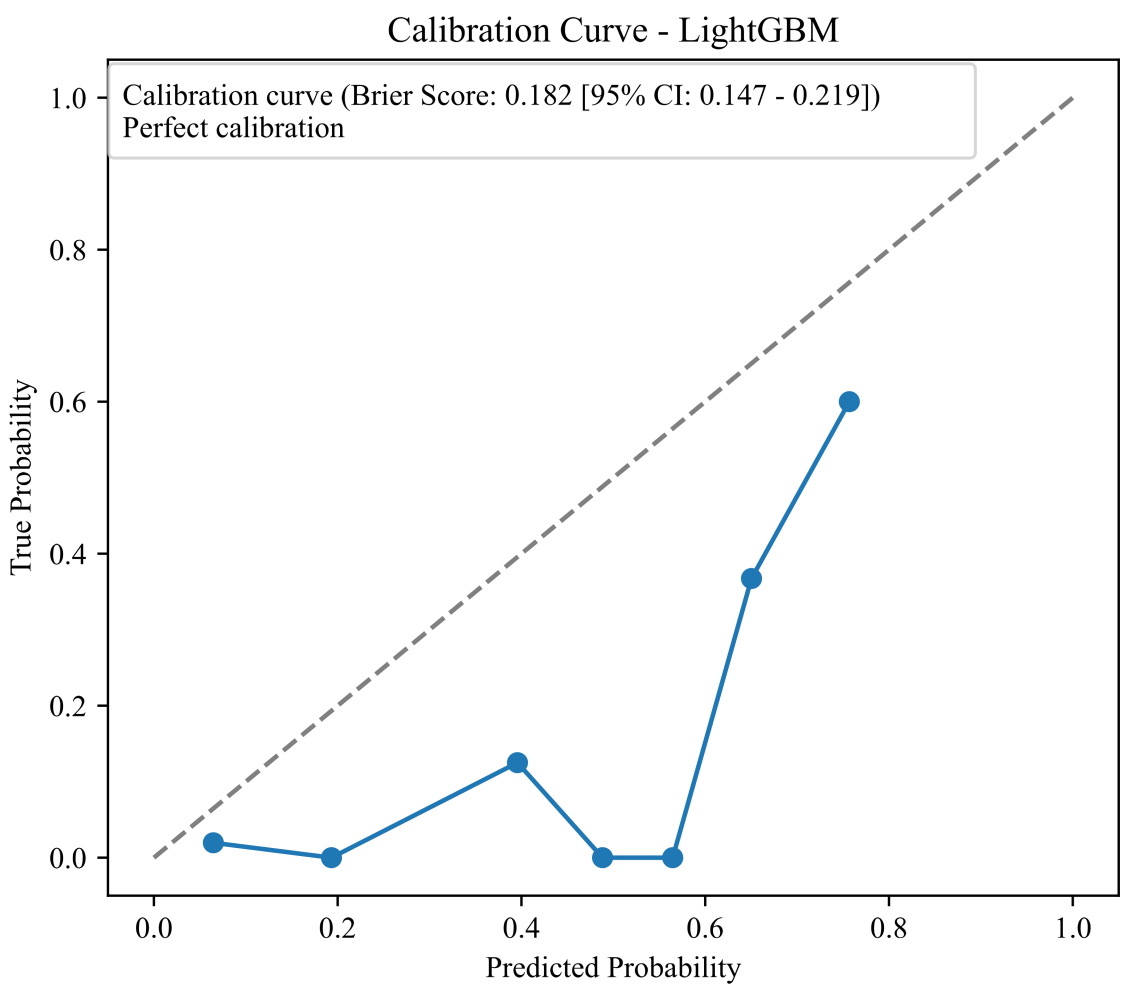
A Training set



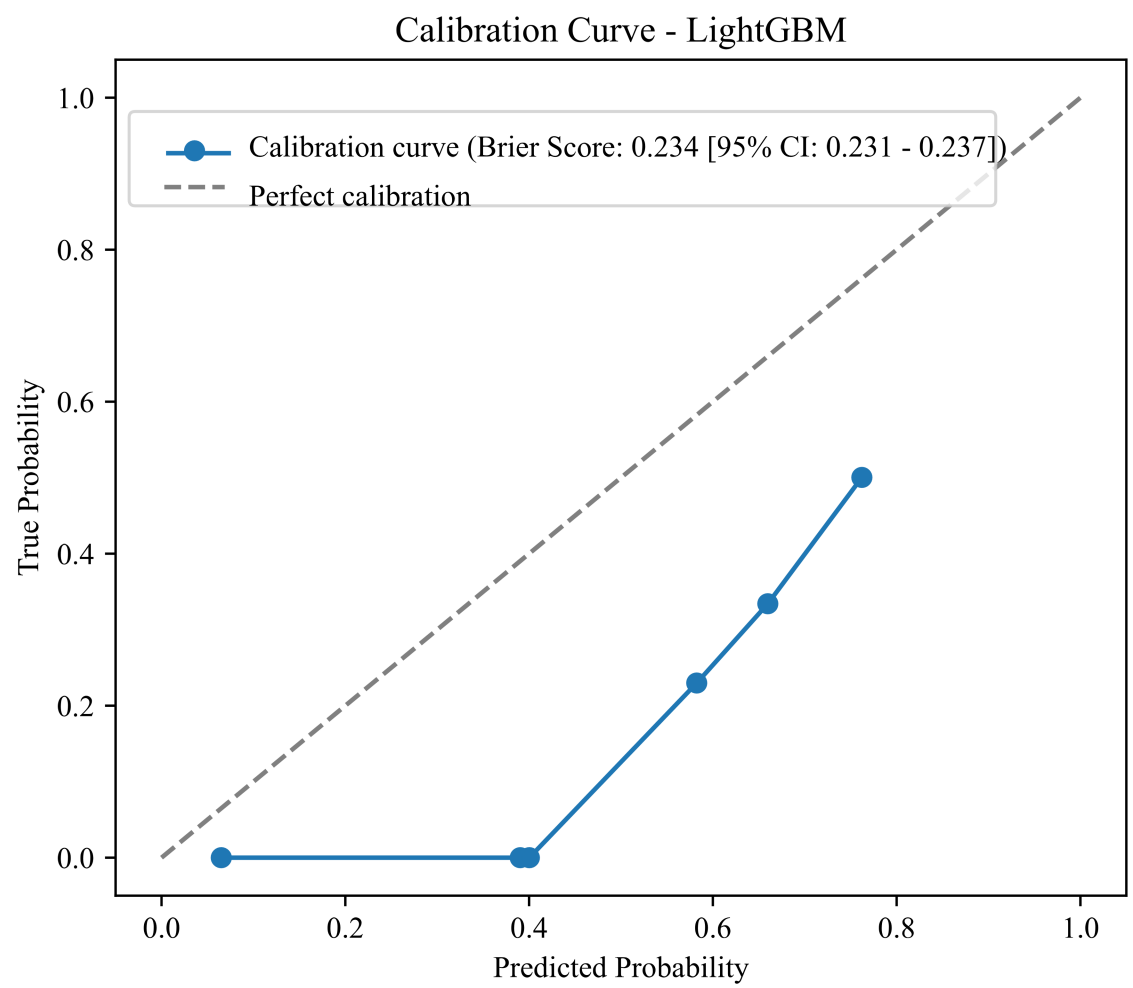
B Internal validation set



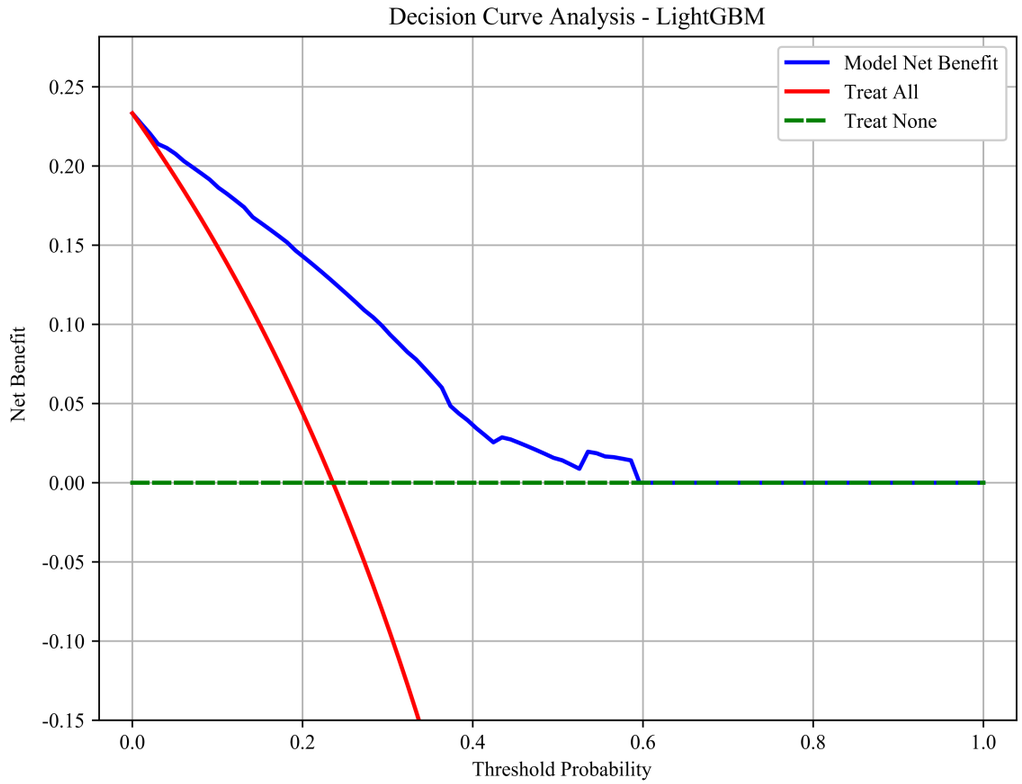
C External cohort 1 (Affiliated Hospital)



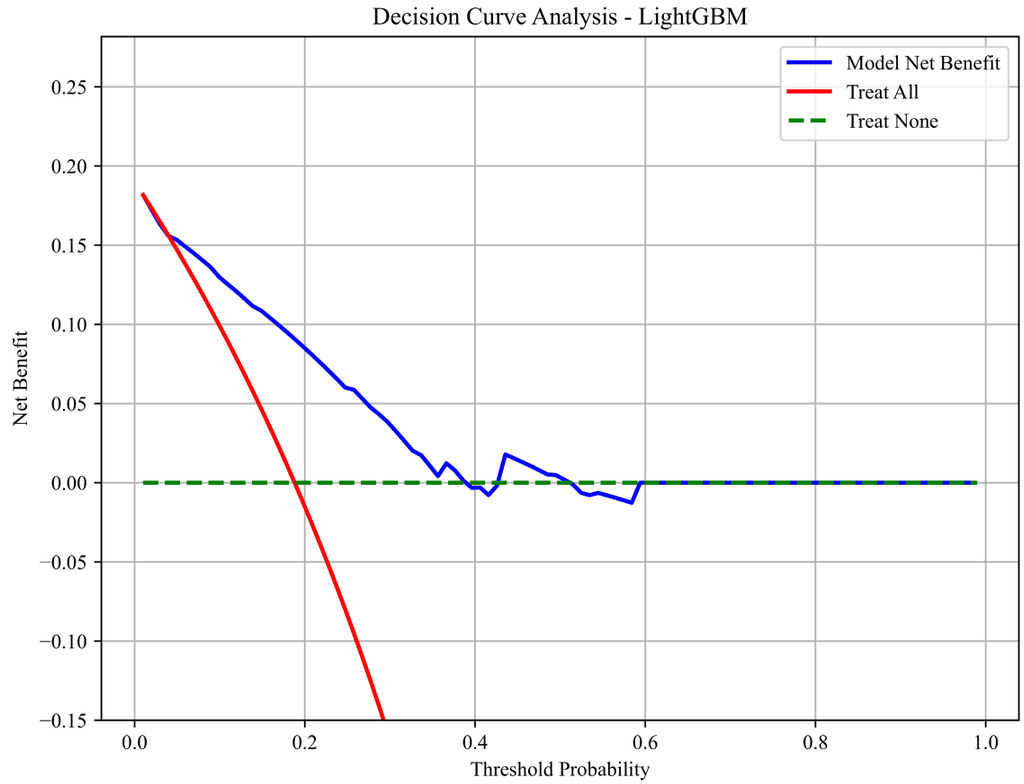
D External cohort 2 (SEER)



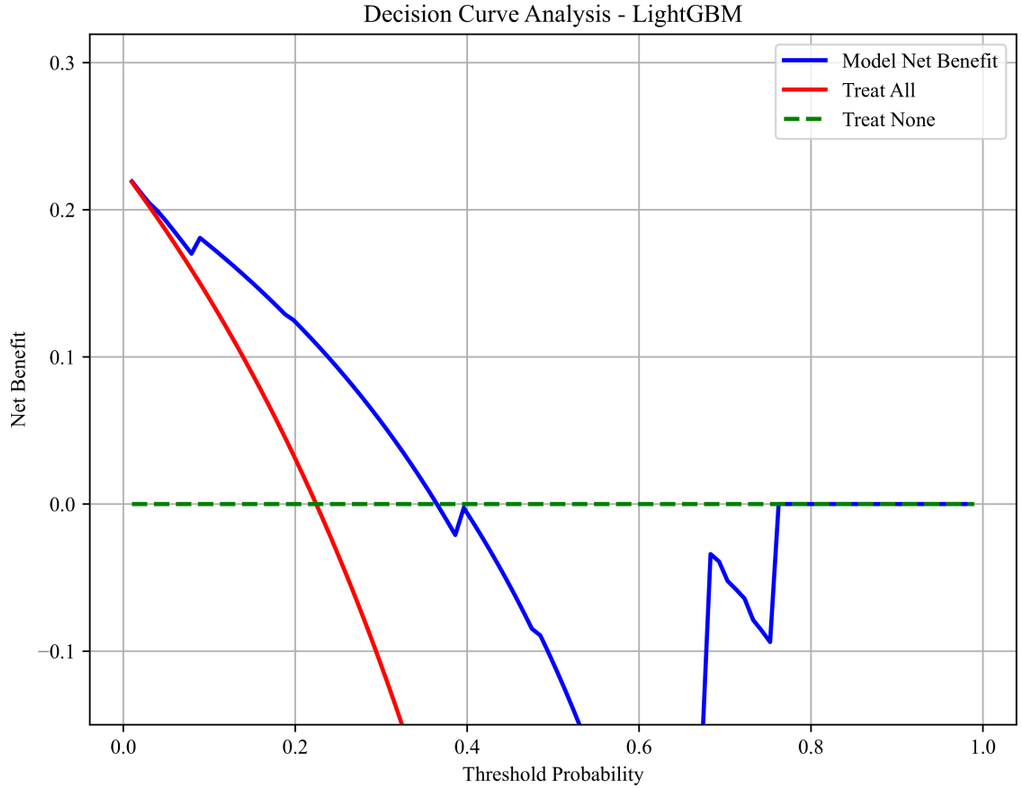
A Training set



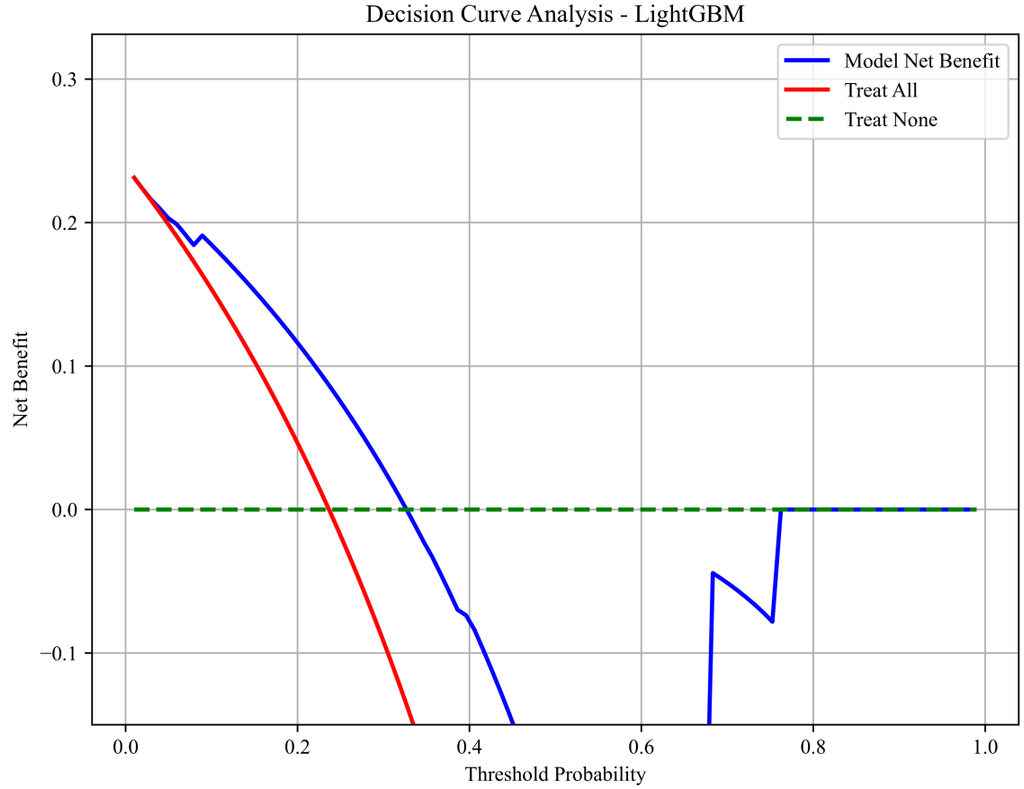
B Internal validation set

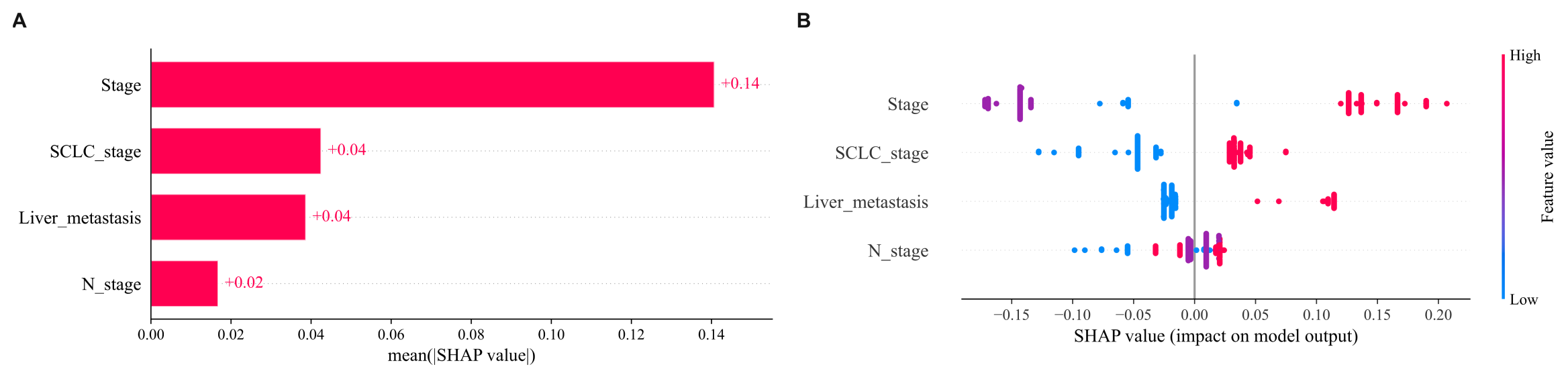


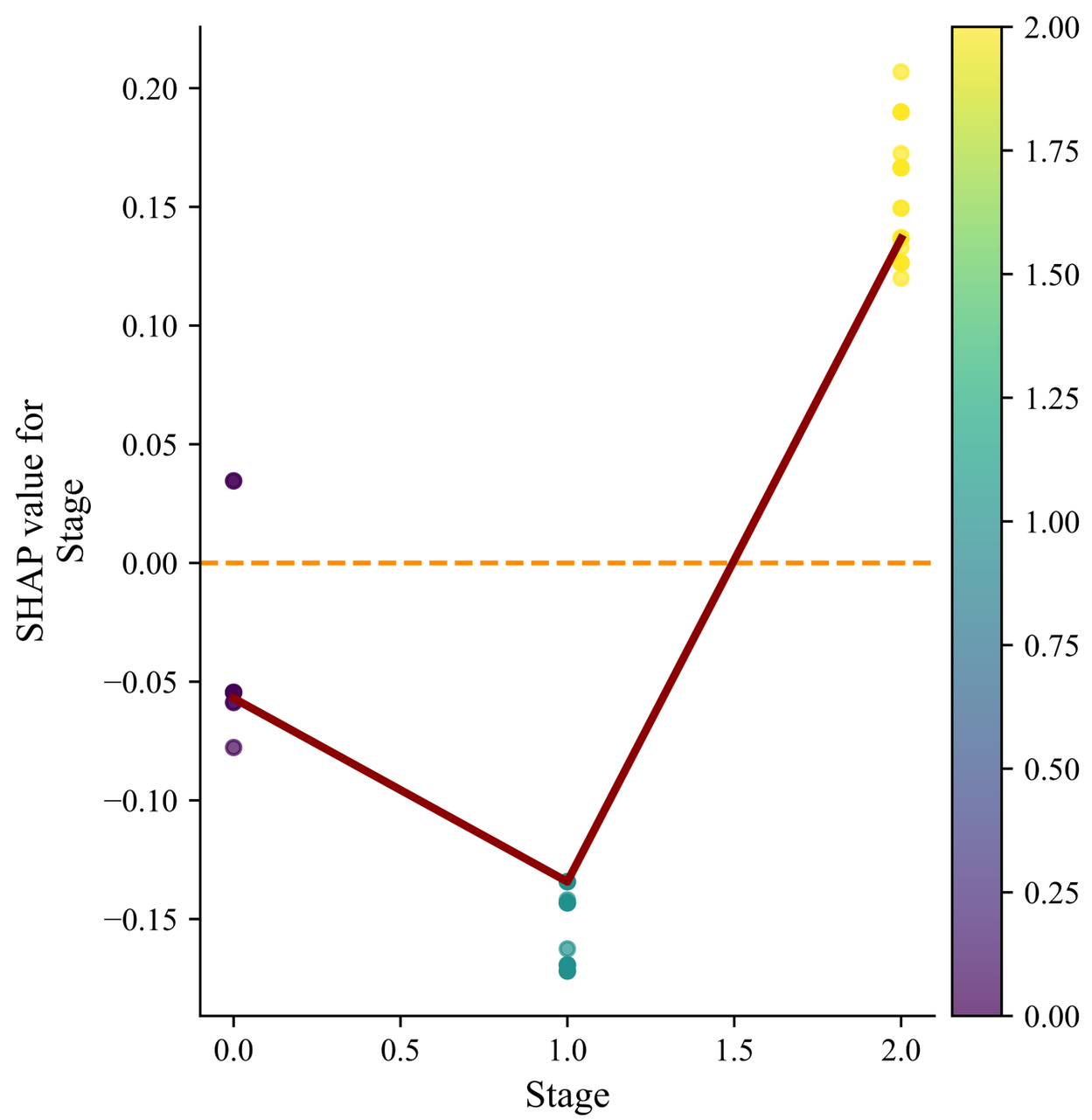
C External cohort 1 (Affiliated Hospital)

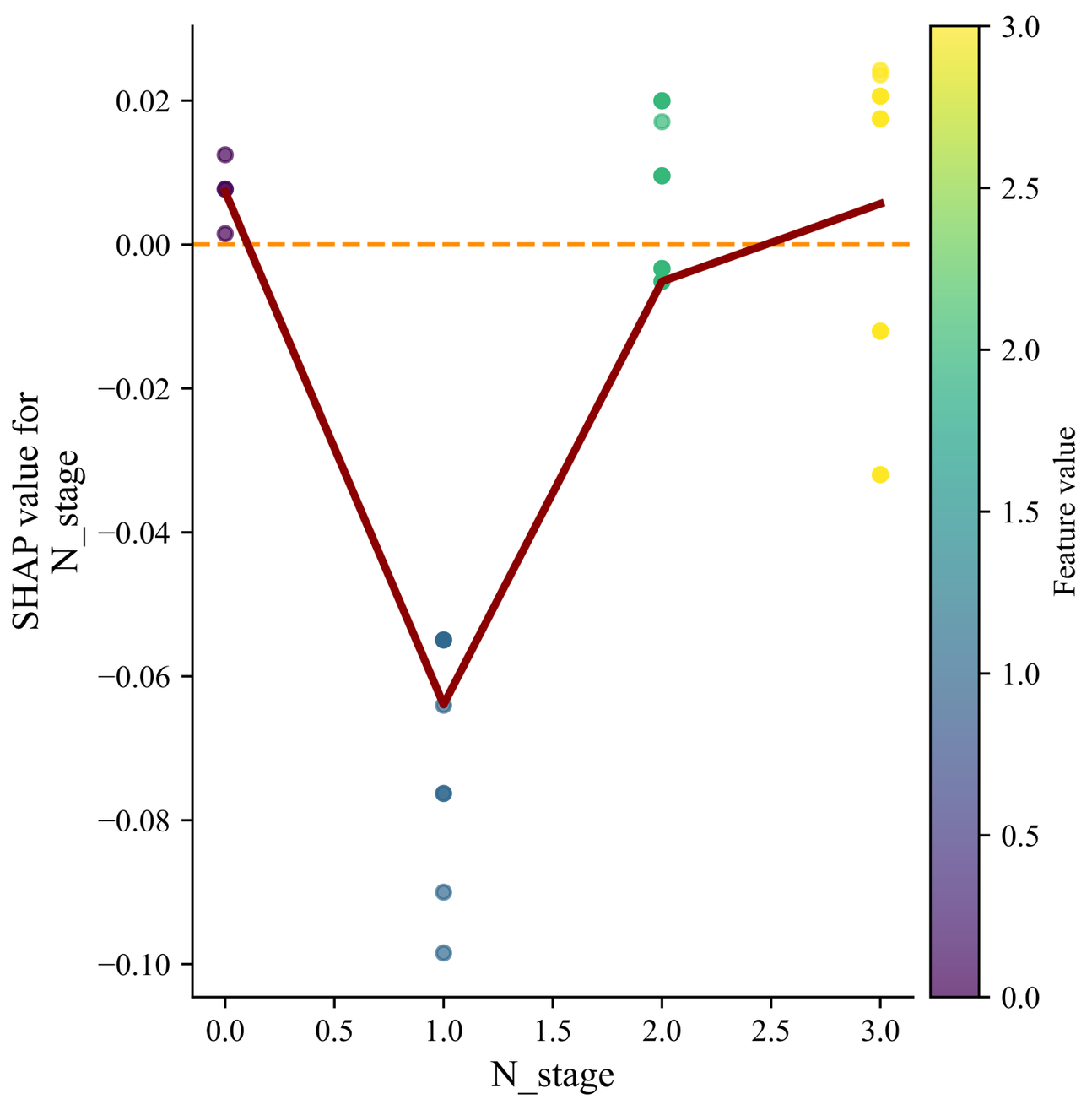


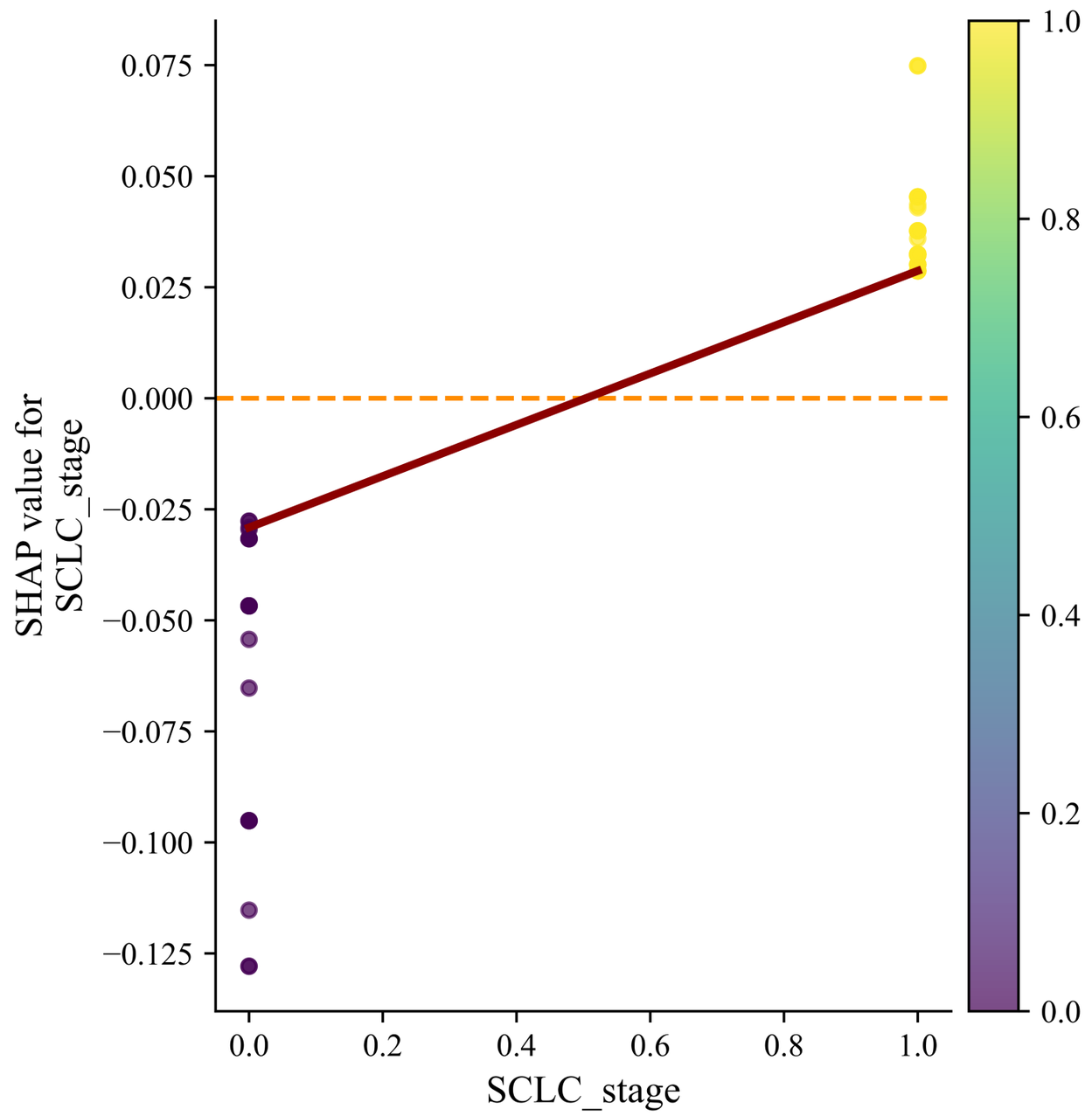
D External cohort 2 (SEER)

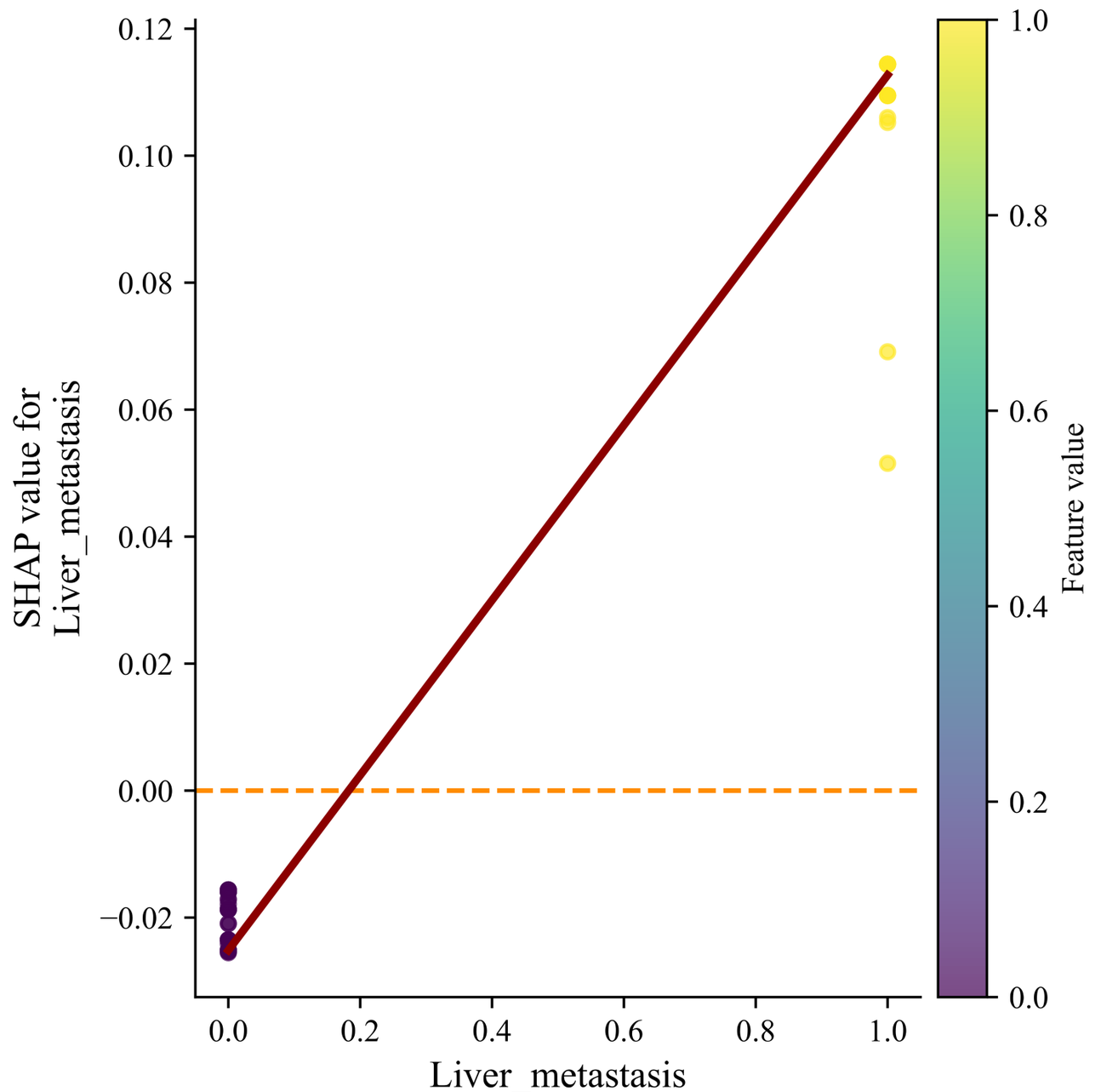


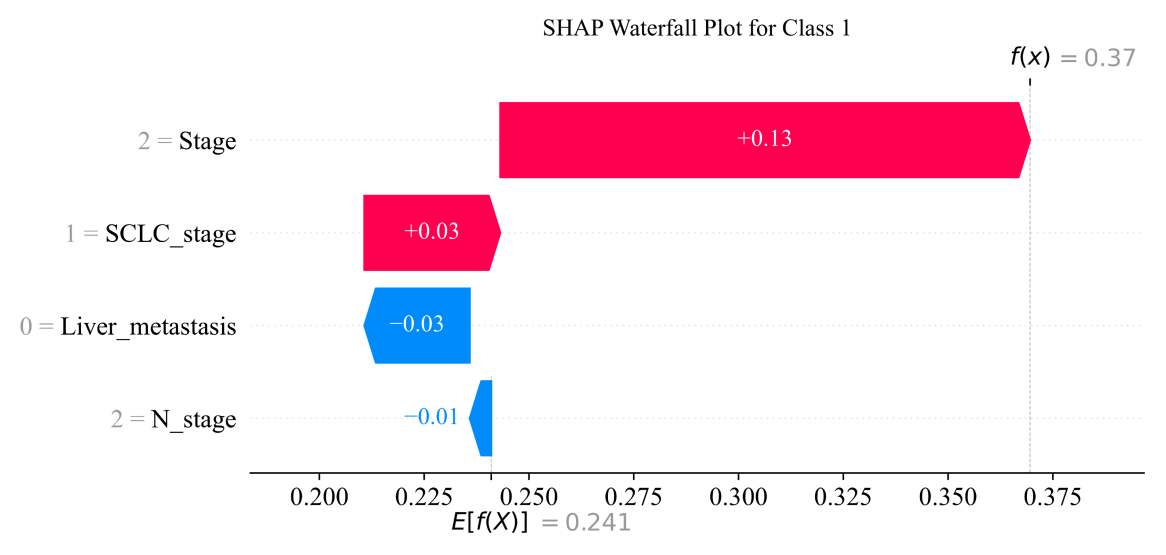

A
Stage
SCLC_stage
Liver_metastasis
N_stage
+0.14
+0.04
+0.04
+0.02
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
mean(|SHAP value|)
B
Stage
SCLC_stage
Liver_metastasis
N_stage
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15
0.20
SHAP value (impact on model output)
High
Low
Feature value


SHAP value for Stage
Stage
Feature value


SHAP value for N_stage
N_stage
Feature value


SHAP value for SCLC_stage
SCLC_stage
Feature value


SHAP value for Liver_metastasis
Liver_metastasis
Feature value


SHAP Waterfall Plot for Class 1
f(x) = 0.37
2 = Stage
+0.13
1 = SCLC_stage
+0.03
0 = Liver_metastasis
−0.03
2 = N_stage
−0.01
0.200
0.225
0.250
0.275
0.300
0.325
0.350
0.375
E[f(X)] = 0.241

Supplementary Files

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

- Additionalfile1TableS1.docx
- Additionalfile2TableS2.docx
- Additionalfile3TableS3.docx
- Additionalfile4TableS4.docx
- Additionalfile5TableS5.docx
- Additionalfile6TableS6.docx
- Additionalfile7TableS7.docx
- Additionalfile8TableS8.docx
- Additionalfile9TableS9.docx
- Additionalfile10SupplementaryMethods.docx