1) **Normalize the continuous variables; dummy-code the categorical variables.**
2) I would make the *final purpose of the model in clinical practice* more explicit. Since you included both variables that are available at the time of diagnosis and variables such as treatment type, pleural effusion, number of lymph nodes, etc., which become available **later**, I would specify that this is a **post-treatment model**. In this way, you avoid criticism regarding a potential **pre-treatment use** of the model.
3) The conclusions of the abstract state *"guide targeted interventions"*, but the use of post-treatment variables somewhat weakens this claim. Check whether we have information on the **treatment sequence**.
4) Filters to apply:

Variable name: Primary Site – labeled

Value	Label
349	C34.9 – Lung, NOS C34.0 Main bronchus C34.1 Upper lobe, lung C34.2 Middle lobe, lung (right lung only) C34.3 Lower lobe, lung

Variable name: Site recode ICD-O-3/WHO 2008

Value(s)	Label
37–41	Respiratory System
39	Lung and Bronchus

Variable name: Behavior code ICD-O-3

Value	Label
3	Malignant

Variable name: Diagnostic Confirmation

Value(s)	Label
1–4	Microscopically confirmed
1	Positive histology
2	Positive exfoliative cytology, no positive histology
3	Positive histology AND immunophenotyping and/or positive genetic studies
4	Positive microscopic confirmation, method not specified

Variable name: ICD-O-3 Hist/behav

Value	Label
2	8000/2: Neoplasm, in situ
3	8000/3: Neoplasm, malignant
7	8001/3: Tumor cells, malignant

11	8002/3: Malignant tumor, small cell type
15	8003/3: Malignant tumor, giant cell type
19	8004/3: Malignant tumor, spindle cell type
23	8005/3: Malignant tumor, clear cell type
42	8010/2: Carcinoma in situ, NOS
43	8010/3: Carcinoma, NOS
47	8011/3: Epithelioma, malignant
50	8012/2: Large cell carcinoma in situ
51	8012/3: Large cell carcinoma, NOS
55	8013/3: Large cell neuroendocrine carcinoma
59	8014/3: Large cell carcinoma with rhabdoid phenotype
83	8020/3: Carcinoma, undifferentiated, NOS
87	8021/3: Carcinoma, anaplastic, NOS
90	8022/2: Pleomorphic carcinoma in situ
91	8022/3: Pleomorphic carcinoma
95	8023/3: Nuclear protein in testis (NUT) associated carcinoma
123	8030/3: Giant cell and spindle cell carcinoma
127	8031/3: Giant cell carcinoma
130	8032/2: Spindle cell carcinoma in situ
131	8032/3: Spindle cell carcinoma, NOS
143	8035/3: Carcinoma with osteoclast-like giant cells
161	8040/1: Tumorlet, NOS
163	8040/3: Tumorlet, malignant
167	8041/3: Small cell carcinoma, NOS
175	8043/3: Small cell carcinoma, fusiform cell
179	8044/3: Small cell carcinoma, intermediate cell
183	8045/3: Combined small cell carcinoma
186	8046/2: Non-small cell carcinoma in situ
187	8046/3: Non-small cell carcinoma
999	8249/3: Atypical carcinoid tumor
1002	8250/2: Bronchiolo-alveolaradenocarcinoma in situ
1003	8250/3: Bronchiolo-alveolaradenocarcinoma, NOS
1007	8251/3: Alveolar adenocarcinoma
1010	8252/2: Bronchiolo-alveolar carcinoma, non-mucinous, in situ
1011	8252/3: Bronchiolo-alveolar carcinoma, non-mucinous
1014	8253/2: Bronchiolo-alveolar carcinoma, mucinous, in situ
1015	8253/3: Bronchiolo-alveolar carcinoma, mucinous
1019	8254/3: Bronchiolo-alveolar carcinoma, mixed mucinous and non-mucinous
1022	8255/2: Adenocarcinoma with mixed subtypes, in situ
1023	8255/3: Adenocarcinoma with mixed subtypes
1027	8256/3: Minimally invasive adenocarcinoma, non-mucinous
1031	8257/3: Minimally invasive adenocarcinoma, mucinous
1040	8260/2: Papillary adenocarcinoma in situ, NOS

1043	8260/3: Papillary adenocarcinoma, NOS
1063	8265/3: Micropapillary adenocarcinoma
923	8230/3: Solid carcinoma, NOS
1923	8480/3: Mucinous adenocarcinoma
1335	8333/3: Fetal adenocarcinoma
579	8144/3: Adenocarcinoma, intestinal type
282	8070/2: Squamous cell carcinoma in situ, NOS
310	8077/2: Squamous intraepithelial neoplasia, grade III
283	8070/3: Squamous cell carcinoma, NOS
287	8071/3: Squamous cell carcinoma, keratinizing, NOS
291	8072/3: Squamous cell carcinoma, large cell, nonkeratinizing, NOS
295	8073/3: Squamous cell carcinoma, small cell, nonkeratinizing
331	8082/3: Lymphoepithelial carcinoma
3891	8972/3: Pulmonary blastoma
3923	8980/3: Carcinosarcoma, NOS
803	8200/3: Adenoid cystic carcinoma
2251	8562/3: Epithelial-myoepithelial carcinoma
1723	8430/3: Mucoepidermoid carcinoma
1243	8310/3: Clear cell adenocarcinoma, NOS
3931	8982/3: Malignant myoepithelioma
962	8240/2: Carcinoid tumor, in situ
963	8240/3: Carcinoid tumor, NOS
987	8246/3: Neuroendocrine carcinoma, NOS
2203	8550/3: Acinar cell carcinoma
2207	8551/3: Acinar cell cystadenocarcinoma

5) The *pleura unknown* variable accounts for **64%**. This likely means that these patients were **not operable**, and they represent a very large group (stage I–II patients who were not candidates for surgery due to poor performance status, comorbidities, advanced age, or alternative treatments). Since these patients have an **unfavorable prognosis**, I am concerned that this could introduce **bias**. I would highlight this either in the **discussion** or in the **limitations** section.

6) I would also make the study limitations more explicit by noting that, since you are using **SEER data**, some **anatomopathological variables** are missing. In the variable selection performed in the multiverse analysis, we had also selected variables such as **lymphovascular invasion (LVI)** and **perineural invasion**. We should evaluate whether these should be included in the model or only discussed in the limitations.

7) Additionally, you report a **63% mortality rate**. Most likely, those who died were **not operated on**. Can this information be extracted? Here too, in my opinion, there is an important **clinical bias**, because it is obvious that non-operated patients have a much worse

prognosis. I would add this to the limitations as well, and possibly include a **table showing whether the deaths occurred predominantly among non-operated patients**.

8) I noticed that you used only **surgery**, **radiotherapy (yes/no)**, and **chemotherapy (yes/no)** variables. In this case, information on the **treatment sequence**—such as chemotherapy followed by surgery or vice versa—is missing. I know that SEER includes some of these variables. If you do not want to add them, I would include a short paragraph discussing the **possible limitations**, because the yes/no approach is somewhat **oversimplified**, and the **temporal sequencing of treatments** clearly has an important prognostic impact.

9) When you write: *"The distant-metastasis component of the AJCC 7th-edition staging system was the single most influential predictor (mean |SHAP| = 37.34). Receipt of surgery (37.03), overall stage group (20.74), nodal status (18.63), and tumor size (13.48) completed the global 'top five'."* I am concerned that the SHAP value of **37** for surgery reflects **sample selection** rather than an intrinsic effect of surgery itself. The *"surgery yes/no"* variable acts as a **proxy for unresectability and clinical severity**, and the high SHAP value likely reflects this selection process rather than a true **biological or therapeutic effect** of surgery per se. I would discuss this explicitly in the **limitations** section.

10) In the waterfall plot, when you report *chemotherapy −12*, I would be cautious with this result, because you are grouping together patients who received **neoadjuvant chemotherapy**, **adjuvant chemotherapy**, and **chemotherapy versus palliative treatment**. Can this information be **disaggregated** / differentiated?