

Investigating Lung-RADS 0: A Retrospective Study Analyzing Clinical Markers and Lung Cancer Risk

Vedika Vyas
vyasvedika05@gmail.com

ABSTRACT

This study investigated Lung-RADS 0 (Lung CT Screening Reporting and Data System) to address gaps in Lung-RADS misclassification and to inform improvements to Lung-RADS 0 (LR0). Clinical markers peripheral capillary oxygen saturation (SpO₂), white blood cell (WBC) count, and serum albumin were examined to determine each's statistical association with lung cancer and used to create a combined logistic regression model to predict binary outcomes for lung cancer. Using a sample of 122 patients from Indiana University, Fisher's exact tests were used to determine associations between each clinical marker using raw data and KNN-adjusted data. It was found in the raw data that having a serum albumin value under 3.5 g/dL was significantly associated with malignancy ($p = 0.0159$). After a secondary sensitivity analysis using KNN adjustment, WBC count became significant ($p = 0.0406$), though the raw data did not support this finding, and serum albumin remained statistically significant ($p = 0.0110$). The logistic regression model produced a cross-validated AUC of 0.53. At the optimal threshold, specificity was 93.9%, but sensitivity was extremely low at 37.5%, which was caused by the lack of positive cases available compared to the negative cases (8 vs 114 cases). Overall, SpO₂, WBC count, and serum albumin alone and combined were not reliable diagnostic indicators of lung cancer in LR0 patients, though larger and more diverse training sets may improve the accuracy of multivariate models, such as logistic regression. Still, research efforts should be redirected towards finding more specific molecular markers or radiological patterns and their association with lung cancer.

INTRODUCTION

Today, lung cancer remains responsible for 20% of all cancer-related mortalities in the United States, the highest number of deaths for any type of cancer, with roughly 124,730 deaths each year (Common Cancer Sites, 2026). With an overall 5-year survival rate of 29.5%, lung cancer remains one of the most aggressive cancers, making it imperative that the disease is diagnosed early (Cancer of the lung and bronchus). One of the main precautionary methods for detecting lung cancer is regular lung cancer screening (LCS) using low-dose computed tomography (LDCT) technology. With updated eligibility criteria from Medicare for free LCSs, the standards are for ages 50-80 years old with at least a 20-pack-year smoking history. This updated criterion has made LCS an option, now more accessible than

ever for early detection. This expanded reach has demanded a rise in LDCT scans performed each year. As a result, radiologists must interpret a growing number of lung nodules. Currently, imaging specialists categorize lung nodules through the American College of Radiology's (ACR) Lung Imaging Reporting and Data System (Lung-RADS): a standardized system classifying lung nodules on a scale from 1, indicating 0 or benign nodules, to 4X, indicating high suspicion of malignancy. As of 2022, the ACR has announced the initiation of the Lung-RADS 0 (LR0) category. This category was meant to act as a group for incomplete or obstructed scans that do not allow a radiologist to accurately interpret the scan. In other words, this subcategory was established to classify patients who needed additional screening because their scans could not be fully evaluated.

However, within Lung-RADS, and particularly LR0, there are high amounts of inter-reader variability amongst scans, especially among non-specialty-trained radiologists compared to cardiothoracic radiologists. At some institutions, it can be found that up to sixty-four to eighty percent of radiologists are accurate, meaning that at least 20% of radiology scans are misclassified (Pavlou et al., 2021). Indicating a need for additional information, clinical markers measured via common serum tests can provide supplemental data for radiologists to use in conjunction with LCS readings, yielding a more accurate reading. This study was conducted through the use of 3rd party websites with the ability to calculate statistics, such as Fisher's Exact Test, and the use of Python code to develop a machine learning model generating a logarithmic odds ratio to input additional values to determine the probability of one developing cancer based on certain clinical marker values. This study aims to determine the statistical significance of the clinical markers via computational modeling and their practical application for radiologists in a clinical setting.

LITERATURE REVIEW

Clinical Markers as a Prognostic Indicator

Clinical marker indices have previously been used to calculate other factors in lung cancer. In fact, other studies have developed inflammation scores as a prognostic biomarker for small cell lung cancer (SCLC), one of the most severe subtypes of lung cancer. In a meta-analysis investigating leukocyte clinical scores and overall survival in SCLC patients, researchers found that patients with a high neutrophil-to-lymphocyte ratio (NLR) have a lower overall survival rate and longer hospital stays than those with a lower NLR (Winther-Larson et al., 2021). This result was further supported by Dr. Füsün Sahin and Dr. Ayşe Feyza Aslan (2024), who also reported higher levels of the neutrophil-to-lymphocyte ratio amongst patients with lung cancer, demonstrating a positive association between NLR and lung cancer development ($p_{\text{trend}} = 0.0001$), indicating high statistical significance. Both studies conclude that, while not definitive, these clinical marker indices can help predict prognosis beyond conventional imaging findings, and could have diagnostic utility when integrated with other clinical data.

Evidence suggests that SpO₂ can be used as a prognostic indicator as well. Martins et al. (2005) investigated pulse oximetry and symptoms in 41 patients with locally advanced or metastatic lung cancer

and found that pulse oximetry was a significant factor in predicting patient prognosis based on survival curve analysis. Patients with lower SpO₂ had significantly poorer survival, with SpO₂ below 95% associated with increased risk of death (OR = 4.02, 95% CI: 1.16–13.85, p = 0.022). The prognostic value shown by Martins et al. (2005) of SpO₂ supports the possibility that SpO₂ may also be useful in identifying lung cancer.

Serum albumin has also been investigated as a prognostic clinical marker in lung cancer. Zhou et al. (2016) demonstrated that the pretreatment albumin-to-globulin ratio (AGR) was a prognostic factor for overall survival in patients with small-cell lung cancer (SCLC), with their findings validated in an independent cohort. Patients with higher AGR values experienced significantly longer overall survival than those with lower values. More recently, a retrospective study of 250 lung cancer patients receiving immunotherapy found that baseline albumin had prognostic value as a predictor of overall survival among patients receiving third-line or later treatment, with higher albumin levels associated with longer overall survival (Chen et al., 2024). Since cancer often causes cachexia, limiting hepatic albumin production, higher albumin levels would indicate a lower severity of cancer-induced cachexia, making serum albumin an inflammatory marker that could have diagnostic value when predicting lung cancer.

Clinical Markers for Early Detection and Risk Prediction

Alternatively, using markers individually has been used to measure lung cancer incidence amongst certain populations. In a study where white blood cell (WBC) count was analyzed for its association with the risk of incidental lung cancer in the United Kingdom, investigators found that the portion of people in the highest quartile of total WBC count had an overall 67% increase in lung cancer risk, an effect caused by neutrophil fractions (Wong et al., 2019). Although this study addresses white blood cell count individually, it fails to use multiple markers in conjunction to make these conclusions, which could overlook a large population of people at risk of developing lung cancer. However, a study by Yoon et al. (2016) investigated risk predictions for lung cancer, calculating risk values for multiple clinical marker values as well as tumor marker values, bridging gaps between marker indices and predictive values by showing how tumor markers and clinical markers can be used in conjunction to diagnose lung cancer. This study ultimately used 12 of the 21 biomarkers investigated to create a predictive logistic regression model, deriving coefficients for each marker from a computational algorithm. Using ROC curve analysis on training and testing data, Yoon et al. (2016)'s model was extremely accurate with an AUC of 0.986.

Mathematical Prediction Models and the Diagnostic Gap in Screening

While it is important to develop statistical models that predict the odds of lung cancer, the applicability of these results in the clinical setting remains unknown. Dozens of calculators intended for radiologists' use do not end up getting used in clinical settings, which is why assessing the clinical relevance of these statistics is of utmost importance. While biochemical markers provide a systemic view of the patient's risk, radiologists often utilize Mathematical Prediction Models (MPMs) to translate imaging findings into actionable percentages. These models, such as the Mayo Clinic Model and the Brock University Model (PanCan model), use logistic regression to calculate the probability of malignancy in cases presented to them. Specifically, the PanCan model has achieved an AUC of 0.84-0.92 (McWilliams et al., 2013), the

highest among all statistical models so far. Despite the precision of mathematical models, the National Lung Screening Trial (NLST) reported a staggering 96.4% false-positive rate for nodules detected via low-dose CT (Chudgar et al., 2011). This highlights a significant "diagnostic gap" where clinical markers could serve as a vital adjunct, as most of these models have inputs of age, smoking history, nodule volume, etc., without any biomarker data. If studies integrate molecular markers with these MPMs, the combined risk profile may shift. Due to the low specificity of current models, if clinical marker data were to be integrated into these models, this could have the potential to reduce the false positive rate.

While all of the aforementioned articles address specific questions, the central component that differentiates this research is the analysis of common clinical markers, including peripheral capillary oxygen saturation (SpO₂), white blood cell (WBC) count, and serum albumin, to develop association values as well as logistic regression models to predict lung cancer risk.

METHODS

Design, Approach

This study observes the effectiveness and applicability of using clinical markers (serum albumin, SpO₂, and white blood cell (WBC) count) for patients whose LCSs are categorized into LR0. The goal of this study was to determine if certain clinical markers or test results had a measurable and significant impact on the diagnosis of lung cancer.

This study should be able to be used in a larger conversation to help improve the gap in diagnostic accuracy. In Pavlou et al. (2021), roughly sixty-four to eighty percent of radiologists read a screening accurately, meaning that it is possible that up to thirty-six to twenty percent of screenings are inaccurate. In fact, in this research, it has been found that LR0, and arguably the entire system, lacks definite criteria, as twenty-seven percent of LDCT LCSs were given a Lung-RADS classification that was not agreed upon by another practicing radiologist.

This study examined a cohort of patients at IU Health who received an LR0 in 2024 and reviewed available clinical marker tests to identify associations between certain results and lung cancer. My exploratory approach allows me to observe patterns and overall trends of certain clinical markers/marker indices and the associated diagnostic outcomes of LR0 patients. This approach has been employed in other areas of research regarding radiology and imaging sciences. These studies typically include a large cohort of people who are studied. Then, results are found by exploring risk factors or markers within these subgroups. However, this study's cohort has been assessed using clinical marker results/levels rather than other factors examined, usually gender and/or smoking status. It is thought that there may be several or combinations of clinical markers (clinical index) that may hold diagnostic value in detecting lung cancer in the future. Zhang et al. (2026) found that an index called CALLY was found to be a prognostic indicator to determine patient outcomes and quality of life. This study influenced the hypothesis as it integrated multiple markers, such as peripheral oxygen saturation (SpO₂), white blood cell count, and

serum albumin, to determine prognostic outlook. Although this study did not look at diagnosis, it certainly influenced how markers should be analyzed.

Methodology/Variables/Selection Criteria

Upon institutional review board approval from Indiana University (IU), this HIPAA-compliant study used a correlational method with an exploratory approach, as the goal of this study was to examine linkages between clinical markers and lung cancer. This study used data from previous patients who received scans from January 1, 2024, to December 31, 2024. Clinical marker data, patient history, biopsy reports, radiological features, and agreement (whether another radiologist agrees with the Lung-RADS classification) were all measured.

Studies that aided in the chosen methodology were Mendoza et al. (2021) and Arora et al. (2025). The paper written by Mendoza et al. (2021) shared characteristics and method approaches that this study chose, in which Mendoza et al. (2021) used radiological features, classification, and patient type data that underwent multivariate analysis, and odds ratios were calculated to determine what risks could be associated with malignancy. This methodology and analysis have similarly been applied to this study, with clinical markers undergoing multivariate analysis, and have been used in an established LCS program. The study completed by Arora et al. (2025) evaluated the clinical use of LR0. Although the aim of the study looked at radiological features and outcomes of LR0 patients, many aspects of this research were shared with the methodology of Arora et al. (2025)'s research, as the cohort classification system of viable lung cancer patients was similar, confirmed by diagnoses from professionals and pathology reports. However, these data will not be used to test the practicality of LR0 in patient settings, but rather to test clinical markers within the LR0 population. Overall, techniques, groupings, and methods were borrowed from each study to fit the needs of this study.

The selection criteria in this study included patients who had a scan between January 1st, 2024, and December 31, 2024, and had available clinical data. To determine the final cohort, LR0 cases from IU Health were selected from a dataset of all the 2024 LCS screenings. Then, as an additional measure, each LCS was verified for accuracy to ensure there were no inconsistencies in the database and establish true LR0 categorization. Since LR0 is an inconclusive classification, a patient's true cancer status could only be established through a subsequent follow-up scan or pathology result; follow-up was therefore required for inclusion so that malignant/benign outcomes could be determined for analysis. From this new cohort, future follow-ups that were taken in any year from the initial screening were examined, where patients were excluded from the study if they did not receive follow-up after the original screening, if their final follow-up screening also resulted in a score of 0, or if the case represented a repeat follow-up for the same patient in 2024.

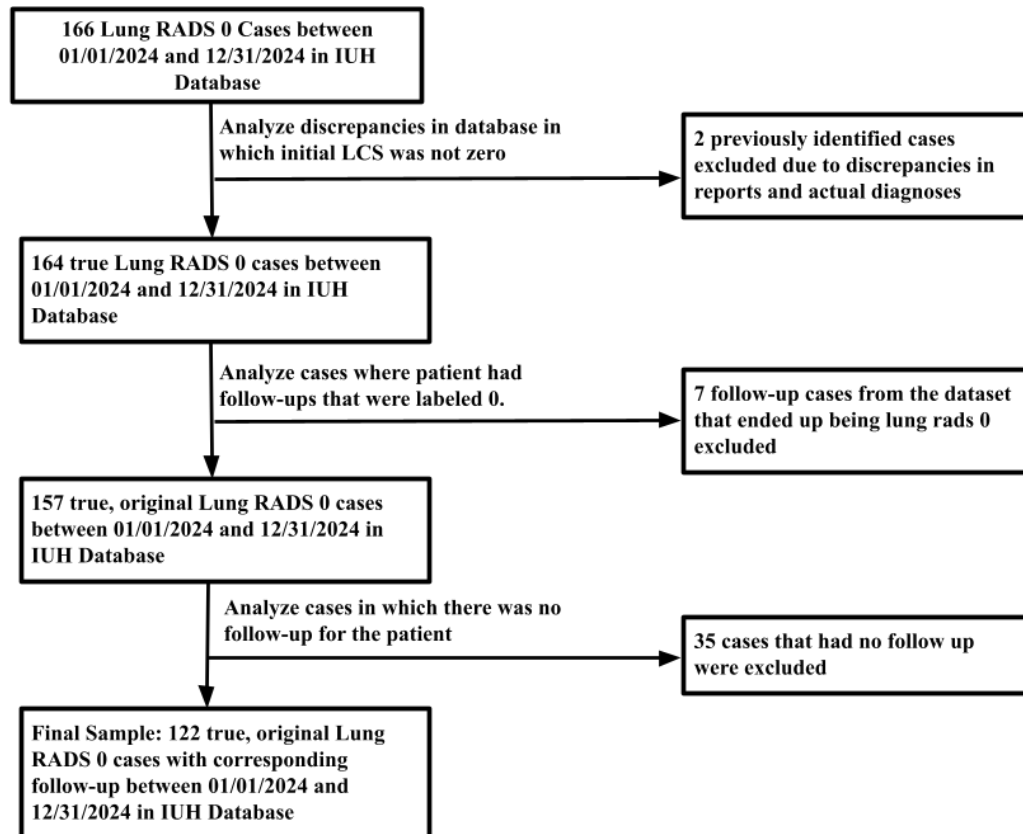


Figure 1: Flowchart for exclusion criteria in the cohort of LR0 patients.

After stratifying the cohort, we checked patient MRNs with a separate dataset of clinical markers (SpO₂, WBC count, and serum albumin) measured within 30 days of the initial screening. Missing values were assigned using a K-Nearest Neighbors (KNN) imputer for secondary sensitivity analysis.

To prevent data leakage and overly optimistic estimates of model performance, the cohort was divided into five folds of 24 - 25 patients each, with the malignant and benign cases distributed as evenly as possible across the folds. Because only eight patients were classified as malignant, each fold had approximately one or two malignant cases, while the rest were benign. For each fold, using only the training data before developing predictions for the unseen validation fold, feature scaling, KNN imputation, and multivariable logistic regression were performed. This process was repeated five times so that each patient was used only once for validation to preserve the integrity of the model. Discriminatory ability was assessed using the out-of-fold Receiver Operating Characteristic (ROC) area under the curve (AUC). The optimal decision threshold was determined using Youden's Index (J). Categorical comparisons using biologically relevant thresholds were evaluated using two-tailed Fisher's Exact tests.

Thresholds for SpO₂ and WBC count were selected based on established clinical cutoffs commonly used to flag physiological abnormality. SpO₂ <95% was used as it is indicative of reduced oxygen saturation;

WBC count >9.0 K/ μ L is the upper limit of the typical reference range, consistent with leukocytosis. For serum albumin, <3.5 g/dL was tested as the threshold to reflect the lower boundary of the normal range (3.5–5.0 g/dL), consistent with hypoalbuminemia. This threshold was selected because low albumin is a recognized marker of chronic inflammation and catabolic states, both of which are commonly observed in malignancy.

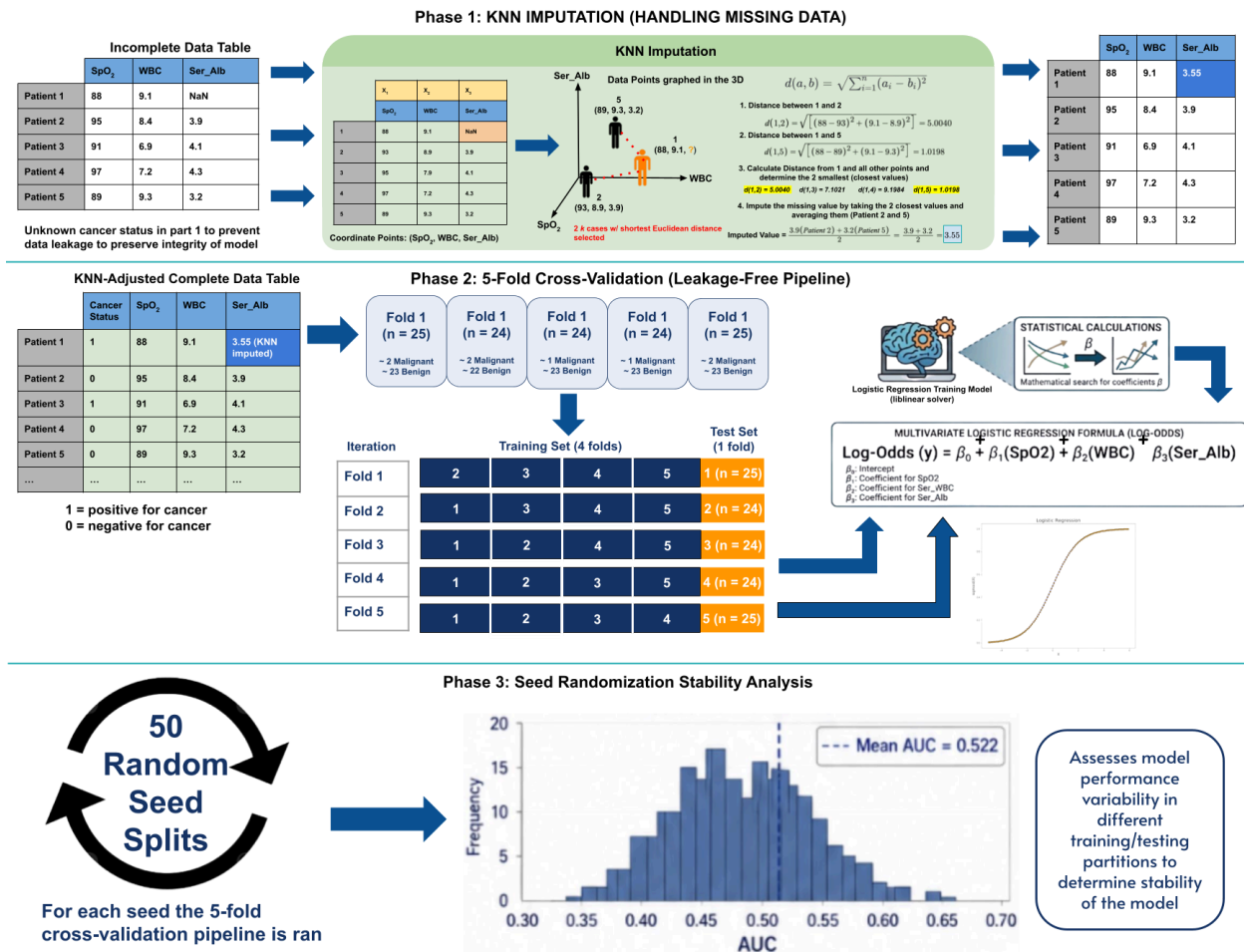


Figure 2: Model of how the KNN imputer replaces missing values of data and is input into a computer program to generate an accurate log-odds formula.

While accounting for similarity between highly variable samples in the data, localized data patterns were emphasized using a neighbor parameter of $k = 2$. To determine whether this selection was appropriate to maintain model discrimination, the complete cross-validated pipeline was repeated using $k = 1$ through 5 to determine whether the number of neighbors calculated during the KNN imputation caused significant differences in the resulting ROC-AUC. After testing k values from 1-5, out-of-fold AUC values ranged narrowly from 0.526 to 0.554 (Figure 3) across this range, indicating that model performance was not meaningfully dependent on the specific number of neighbors used for KNN imputation. Due to the small

number of malignancies, differences in AUC across $k = 1$ to 5 likely reflect sampling noise rather than actual improvements in imputation quality. Selecting k post hoc on which k produced the highest AUC would risk overfitting the imputation hyperparameter to this noise, creating an overoptimistic prediction of model performance. Accordingly, $k = 2$ was retained as the pre-specified value. Additionally, continuous variables were adjusted using z-score normalization to ensure distance calculations were not skewed by differences in measurement scales. To further assess the stability of model performance outside of k , the model (fixed at $k = 2$) was repeated across 50 different random folds.

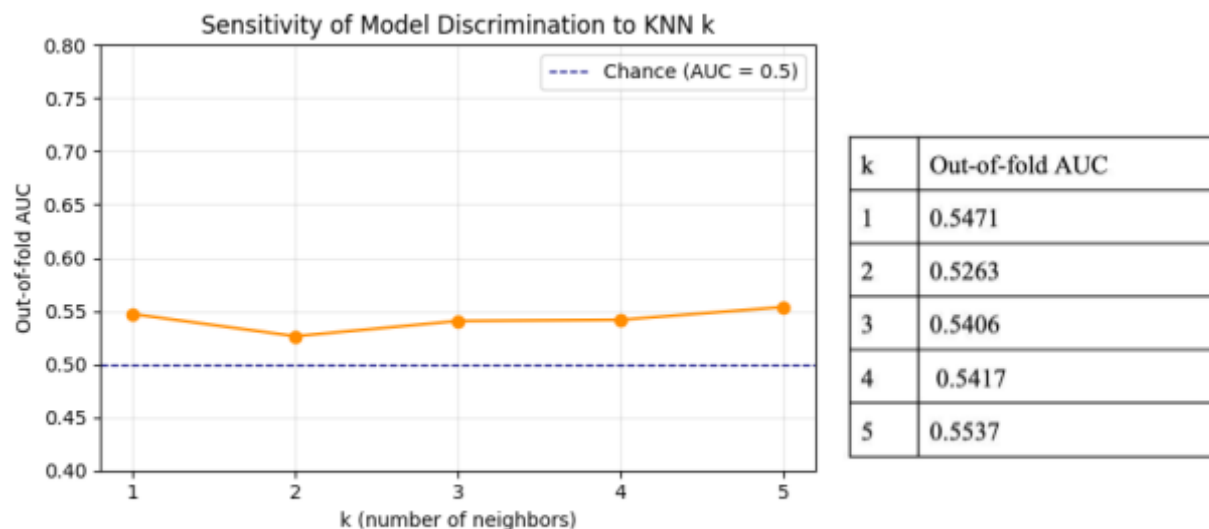


Figure 3: Out-of-Fold AUC Across KNN Neighbor Values (k)

Out-of-fold AUC across these 50 fold assignments ranged from 0.355 to 0.662 (mean = 0.522, SD = 0.046; 95% range: 0.453 - 0.617), with the fold used for the figure (AUC = 0.526) falling close to this distribution's mean (Figure 4). This is relatively wide compared to the narrow midpoint value, which shows that, while only 8/122 cases were malignant, the out-of-fold AUC is highly sensitive to the different fold groupings, likely because of the small number of confirmed malignancies. Accordingly, the single-seed AUC estimates in this study indicate that the discriminatory capability of this model is similar to that of a coin flip, rather than precise, reproducible values.

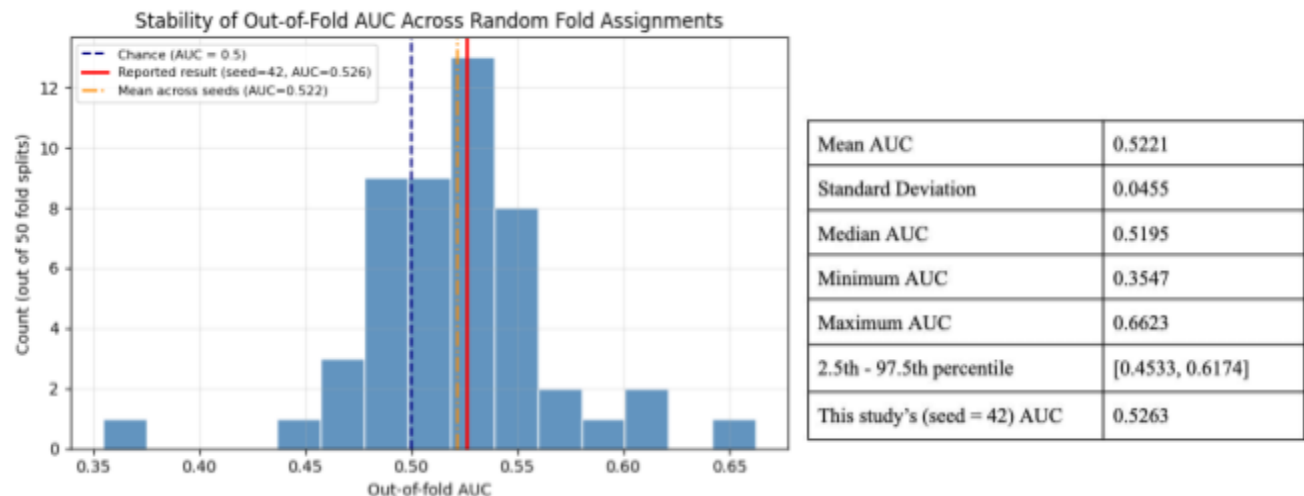


Figure 4: Out-of-Fold AUC Stability Across 50 Random Fold Splits (k=2)

Materials/Procedure

IU personnel retrieved the required information needed to conduct the study. Information regarding the initial screening result, follow-up screenings, and clinical lab data were all sent in order to find members of the final cohort studied. Then, based on selection criteria, each scan was included or excluded from the final cohort. After determining which patients had available results for the clinical tests, those values were analyzed via Google Colab and Excel to identify associations between specific markers and develop multivariate statistical models to assess lung cancer risk.

Limitations/Anticipated Analytics/Ethics

Although this study strived for the highest caliber of accuracy and applicability, several methodological limitations must be acknowledged. First, the study took a retrospective approach; such causal relationships could not be established between clinical markers and outcomes. Secondly, the dataset was substantially incomplete among clinical markers before imputation, with complete raw data available for only 65 patients for SpO₂, 30 for WBC count, and 28 for serum albumin. Z-score standardization and KNN imputation were used to retain sample size (n = 122) and preserve covariance across variables. However, reliance on these methods while data was sparse introduced estimation bias that could have been limited if the sample was more robust. Most notably, the cohort contained a low number of patients who developed lung cancer, with only 8 of the 122 patients having a confirmed diagnosis via pathology. While a 5-fold stratified cross-validation tried to prevent data leakage and minimize optimistic outputs, fitting a three-variable logistic regression model on a small number of events created major statistical power constraints and instability within the model. Specifically, the low events-per-variable ratio can increase standard errors, leading to wide confidence intervals for regression coefficients and odds ratios of each marker. This instability can also affect the model's discriminatory capabilities. Because the model is based on a small number of positive events, it's more likely to overfit to those cases, causing the "apparent" AUC to be artificially high. However, when tested using out-of-sample cross-validation, the

model's ability to predict risk may decrease, resulting in a lower, more realistic out-of-fold AUC. The small sample size can also push predicted probabilities toward more extreme values, making diagnostic performance sensitive to small changes in each classification. This causes sensitivity to be particularly unstable, as misclassifying one of the few positive cases can cause an outstanding change in the sensitivity estimate. On the other hand, specificity may remain artificially high because the dataset contains a larger number of benign cases, making the model appear more effective at correctly identifying patients without cancer. Therefore, the out-of-fold AUC, sensitivity, and specificity must be interpreted with caution, as small fluctuations within the sample can produce unstable or overly optimistic estimates of model performance.

These findings reflect the preliminary, exploratory nature of this study rather than a validated clinical tool. Fully establishing the diagnostic utility and generalizability of these clinical markers will require prospective, multi-institutional studies with substantially larger sample sizes and rigorous external validation datasets.

Despite these constraints, this methodology was chosen because it provided access to consistent clinical records, aligned with the study's exploratory goals, and offered a practical, ethically sound way to evaluate trends in LR0 patients without overstepping available data.

Ethical considerations were addressed during this experiment. This experiment required an application for IRB approval from Indiana University, in which HIPAA compliance was a requirement for the approval of this study. Patient consent was not required due to the retrospective nature of the study. Patient information was stored digitally in encrypted files and password-protected devices, ensuring that patient data was kept safe and privacy was respected. After the experiment had been concluded, all data was destroyed/deleted.

RESULTS

The study conducted a comprehensive analysis of 122 patients from IU University Hospital to investigate the association between clinical markers and lung cancer outcomes, specifically within the LR0 category. 8 out of the 122 LR0 patients were found to have lung cancer through confirmation of pathology reports. The remaining 114 patients had benign nodules. These clinical tests were for SpO₂, WBC count, and serum albumin. We performed statistical analyses using Google Colab and Excel. Google Colab was chosen as the primary coding platform to conduct statistical analyses due to its vast set of data science libraries, simple setup with pre-installed libraries, and affordability. By reviewing patient scans alongside available clinical marker data, the research aimed to determine if these markers could reliably predict diagnostic outcomes and help reduce inter-reader variability among radiologists. The primary research objective was to assess the discriminative ability of the aforementioned marker index to identify cancerous manifestations in the lungs.

P-values of Raw Marker Values Using Fisher's Exact Test

Fisher's Exact Test was performed on the raw marker values for SpO₂, WBC count, and serum albumin. Thresholds for each marker were determined by finding clinically relevant cutoffs for each marker based on established medical guidelines or values commonly associated with patient risk; for example, an SpO₂ under 95% is considered concerning. Similarly, thresholds for WBC count and serum albumin were categorized accordingly. The corresponding two-tailed p-values for SpO₂ < 95%, WBC count > 9.0 K/ μ L, and serum albumin < 3.5 g/dL were 0.223515, 0.326828, and 0.0159, respectively (Figure 5). Only the threshold of serum albumin being under 3.5 g/dL seemed to have a statistically significant result ($p = 0.0159$). In a prospective study run by Yang et al. (2021) studying serum albumin and cancer outcomes, although Yang et al. (2021) found an inverse relationship between serum albumin levels and overall cancer risk, it was also found that associations between albumin and lung cancer were nonsignificant, aligning with the results of this study.

Cross-tabulation of Raw Marker Values and Diagnostic Outcomes								
Clinical Marker Raw Data			Diagnostic Outcome		Total	Two-Tailed P Value	95% CI	Odds Ratio
			Malignant	Benign				
	SpO ₂ (%)	<95	4	23	27	0.223515	[0.5299, 18.4929]	3.1304
		≥ 95	2	36	38			
		Total	6	59	65			
	WBC (K/ μ L)	>9.0	3	8	11	0.326828	[0.4415, 23.0112]	3.1875
		≤ 9.0	2	17	19			
		Total	5	25	30			
	Serum Albumin (g/dL)	<3.5	2	0	2	0.0159	[1.8009, 1333.2111]	inf.
		≥ 3.5	2	24	26			
		Total	4	24	28			

Figure 5: P-values, confidence intervals, and odds ratios of SpO₂, WBC (mL), and Serum Albumin (g/dL) of LR0 patients.

Sensitivity Analysis of KNN-Adjusted Marker Values Using Fisher's Exact Test

As a secondary sensitivity analysis, missing data were imputed using KNN, and associations were evaluated using Fisher's exact test. The two-tailed p-values for SpO₂ < 95%, WBC count > 9.0 K/ μ L, and serum albumin < 3.5 g/dL were 0.4731, 0.0406, and 0.431, respectively (Figure 6). A statistically significant association was found for a WBC count over 9.0 K/ μ L and lung cancer ($p = 0.0406$) and serum albumin under 3.5 and lung cancer ($p = 0.0110$). While the statistically significant relationship between serum albumin and lung cancer was observed in the raw data, this relationship was not observed for a WBC count over 9.0 K/ μ L in primary raw data, suggesting that the finding correlating to WBC count may reflect imputation-related biases rather than a true effect, as missing values are estimated by averaging/weighting values from the most similar observations in the dataset, which can amplify

associations absent from the dataset. In a study assessing imputation methods for missing data in predictive modeling of cohort studies, it was found that although KNN was the most effective method of data imputation out of all the tested models, there were still discrepancies within KNN's accuracy (AUC = 0.769), which may account for this inconsistency (Li et al., 2023). However, for markers excluding WBC count, insignificance was maintained across the KNN-adjusted and raw values.

Cross-tabulation of KNN-Adjusted Marker Values and Diagnostic Outcomes								
			Diagnostic Outcome		Total	Two-Tailed P Value	95% CI	Odds Ratio
			Malignant	Benign				
Clinical Marker KNN-Adjusted Data	SpO2 (%)	<95	6	67	73	0.4731	[0.4069, 10.8838]	2.1045
		>=95	2	47	49			
		Total	8	114	122			
	WBC (K/ μ L)	>9.0	3	19	22	0.0406	[1.1486, 21.7649]	5
		<=9.0	5	95	100			
		Total	8	114	122			
	Serum Albumin (g/dL)	<3.5	2	1	3	0.0110	[2.9794, 476.1979]	37.6667
		>=3.5	6	113	119			
		Total	8	114	122			

Figure 6: P-values, confidence intervals, and odds ratios of SpO2, WBC (mL), and Serum Albumin (g/dL) of LR0 patients with KNN-adjustments.

Logistic Regression & ROC Curve Analysis

A logistic regression was developed in Google Colab to predict lung cancer outcomes using a combination of clinical markers to produce a single log-odds equation capable of predicting binary outcomes for lung cancer. The resulting model produced the following equation for the log-odds of lung cancer:

$$\text{Log-Odds: } -2.4300 + -0.0135 * \text{SpO}_2 + -0.3066 * \text{WBC} + -1.0105 * \text{Ser_Alb}$$

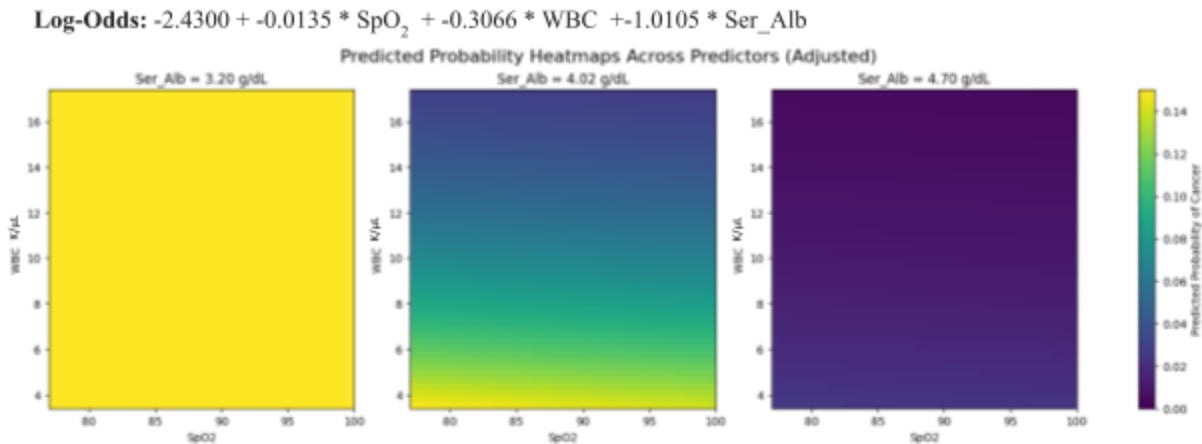


Figure 7: A set of predicted probability heatmaps visualizing the log-odds equation for various serum albumin values. WBC and SpO_2 have continuous scales, whereas Serum Albumin has a fixed scale, in which the leftmost is at the minimum value (3.20 g/dL), the middle at the mean (4.02 g/dL), and the rightmost at the maximum serum albumin value (4.70 g/dL).

Model performance was assessed by way of a receiver operating characteristic (ROC) curve, where the area under the curve (AUC) value of 0.53 shows very poor discriminatory power. The optimum probability cutoff point was found to be 0.19 with the highest Youden's index value of 0.3136. The sensitivity and specificity values for the optimum probability of 0.19 are 0.375 and 0.939, respectively. Such an imbalance implies that the model can successfully identify most of the benign cases as benign but is unable to identify most of the malignant cases, which makes this model a poor diagnostic tool. The high specificity value (0.9649) and low sensitivity value (0.375) of the model suggest the diagnostic inadequacy of this model to either rule out or identify any disease because such a combination is not good for any diagnostic purpose due to low sensitivity. Since AUC shows the performance of the test at all cutoff points, the AUC value of 0.53 suggests that the high specificity at the 0.19 cutoff point is not indicative of diagnostic separability. In a study conducting ROC curve analysis on numerous regression models based on the advanced lung cancer inflammation index (ALI), Hu et al. (2021) found that Youden's index for WBC count and serum albumin individually were 0.531 and 0.250, respectively, with the ALI index overall having a maximum Youden's index of 0.094, with an optimal cutoff value of 48.2 in these models. This indicates that other studies exploring indices with similar markers also experience limited overall diagnostic effectiveness. However, other studies have shown success in developing such predictive models. Zhai et al. (2024) reported AUCs in an ROC curve for albumin and neutrophil-to-leukocyte ratio of 0.977 and 0.796, respectively, suggesting high diagnostic accuracy of both markers. Thus, with further refinement, it may be possible to achieve higher sensitivity with this model.

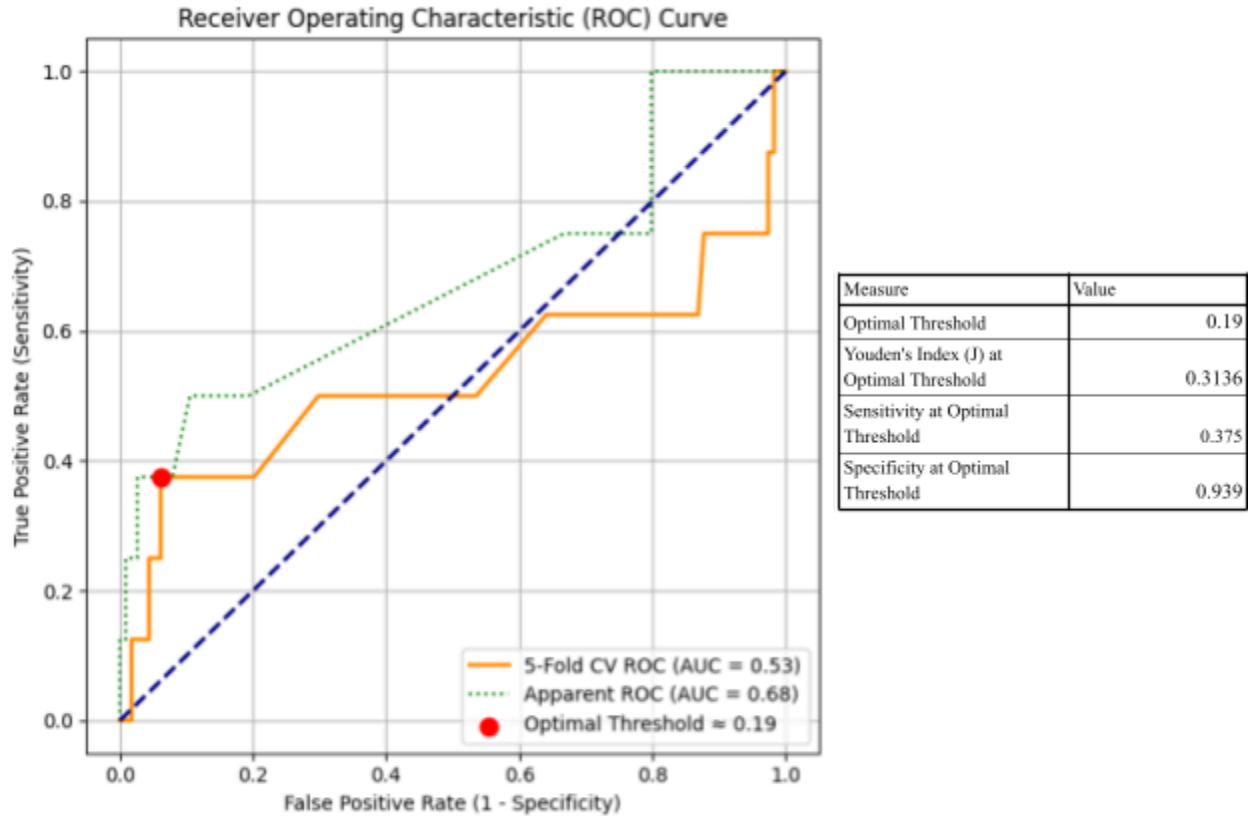


Figure 8: ROC Curve with corresponding values (listed).

Data Distribution

Boxplots configured in Microsoft Excel illustrated the distribution of SpO₂, WBC count, and serum albumin in cancerous and noncancerous patients using the primary, raw data. Given the small number of patients who had cancer, these comparisons have been included to illustrate descriptive trends within this cohort and are not definitive evidence of biological associations. Median values were compared to account for outliers in the data. In SpO₂, cancerous patients were found to have a lower median value of SpO₂, with a difference of -2.5 (Figure 9). Median WBC counts were higher in cancer patients, with a difference of +0.5 K/ μ L (Figure 10). Median serum albumin levels were found to be reduced in cancerous processes compared to benign nodules, with a difference of -0.75 g/dL (Figure 11).

SpO₂ (%) Values of Cancerous and Noncancerous LR0 Patients

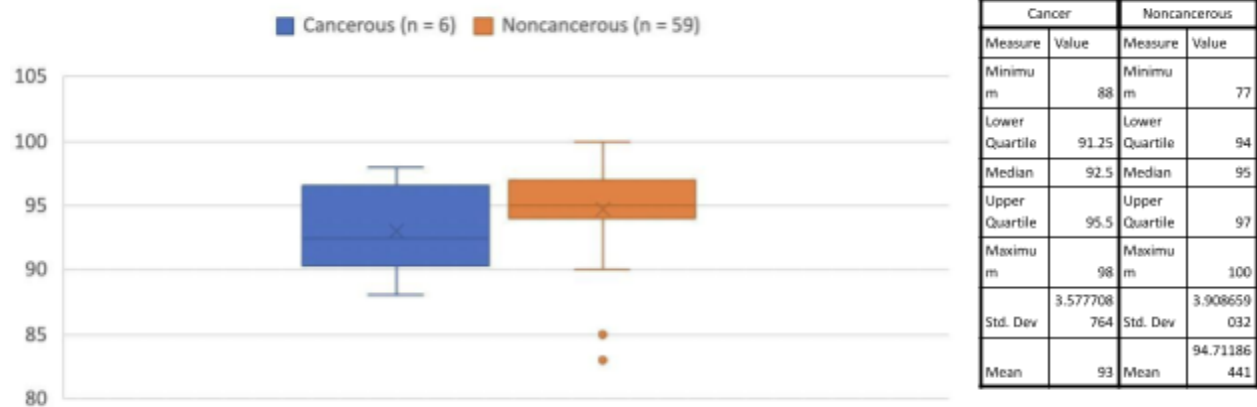


Figure 9: Boxplot representation of the distribution of SpO₂ values in cancerous and noncancerous LR0 patients.

WBC Counts (K/ μ L) of Cancerous and Noncancerous LR0 Patients

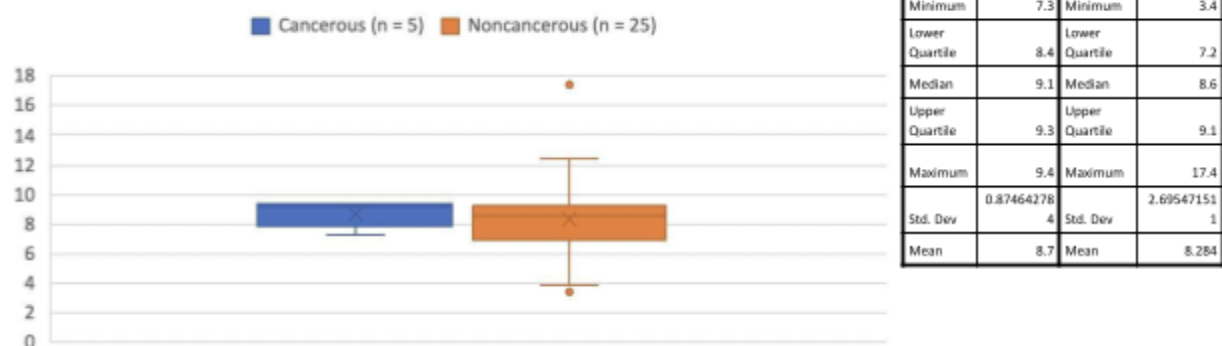


Figure 10: Boxplot representation of the distribution of WBC counts (mL) in cancerous and noncancerous LR0 patients.

Serum Albumin (g/dL) Values of Cancerous and Noncancerous LR0 Patients

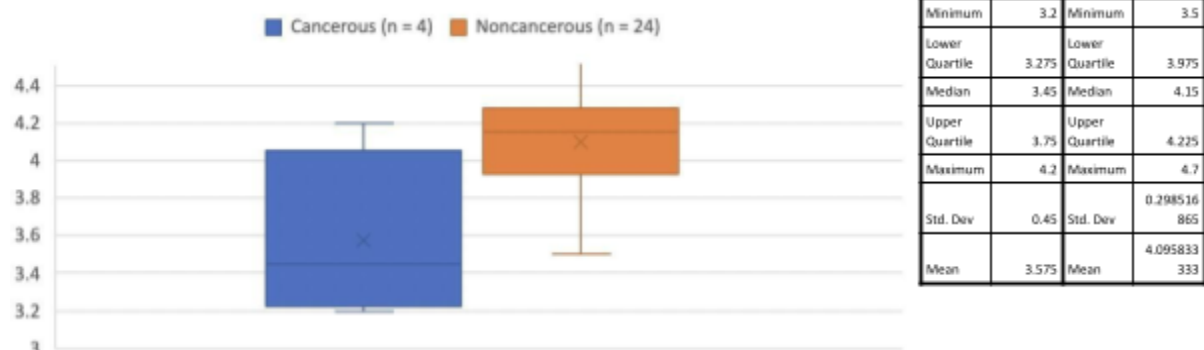


Figure 11: Boxplot representation of the distribution of serum albumin (g/dL) in cancerous and noncancerous LR0 patients.

ANALYSIS

The findings from this study suggest that while serum albumin levels below 3.5 g/dL were statistically significantly associated with lung cancer risk ($p = 0.0110$), routine clinical markers overall provided limited diagnostic utility for LR0 screening. While these markers may not help radiologists discern between LR0 cases, these findings have shed light on how MLMs for risk stratification in LR0 must be refined. By establishing that combining SpO₂, WBC count, and serum albumin offers low diagnostic utility, overreliance on adjunctive, nonspecific clinical indicators while interpreting readings is discouraged. While clinical markers have shown prognostic value in other studies, such as one conducted by Akcam et al. (2025), where they found that non-small cell lung cancer patients with a higher PLR (platelet-to-lymphocyte) ratio, on average, had lower overall survival rates, longer hospital stays, and took longer to be eligible for chest tube removal. Ultimately, using such clinical markers as a diagnostic indicator still remains under-researched.

SpO₂

Although the median SpO₂ value for malignant patients was 92.5%, while the median for benign patients was 95%, there was a statistically insignificant association between SpO₂ and lung cancer malignancy, as the two-tailed p-value for the raw and KNN-adjusted data for SpO₂ values over/under 95% and lung cancer were 0.223515 and 0.4731, respectively (Figures 5 & 6). The coefficient for SpO₂, as per the logistic regression model, was -0.0135, indicating almost no association with cancerous outcomes. This relationship is biologically plausible because SpO₂ is a nonspecific physiological measurement that reflects blood oxygenation rather than tumor presence. Lung cancer, particularly in its early stages, may not impair gas exchange to such a degree that it would alter SpO₂ levels. Decreased SpO₂ is associated with a multitude of other disorders, meaning that decreased SpO₂ could be indicative of conditions unrelated to lung cancer. It should be noted that only 8 malignant cases were present within the dataset, with only 6 of those 8 patients having available clinical data for SpO₂; this sparse dataset may further contribute to the insignificant findings related to this clinical marker.

White Blood Cell (WBC) Count

Statistics based on WBC count had varied results for its associations with lung cancer. While a statistically significant p-value was produced in the KNN-imputed data for WBC count over 9.0 K/ μ L ($p = 0.0406$), in the primary raw data analysis, there was no statistically significant association ($p = 0.3268$). Given that complete WBC count data were available for 30 patients, this sensitivity to imputation indicates that the post-imputation significance occurs because of the distance-weighted imputation on a small dataset, rather than an actual association (Figures 5 and 6). Although the median WBC count for malignant patients was 9.1 and the median WBC count for benign patients was 8.6, only 5 of 8 malignant cases were available for analysis. This lack of information may contribute to imputation and sampling-related biases, causing unforeseen statistical associations when, in reality, no such relationship exists. Data that further supports the conclusion that WBC count levels alone are unassociated with lung cancer is the WBC count coefficient ($\beta = -0.3066$) in the logistic regression model (Figure 7). However, other prospective studies with larger sample sizes have found associations between elevated WBC count and lung cancer risk ($p_{\text{trend}} = 1.4 \times 10^{-12}$), indicating that perhaps with a larger cohort size, a more sizeable

difference may be observed. However, for the scope of this study, no such association existed (Wong et al., 2019).

Serum Albumin

Serum albumin had a statistically significant association with malignancy at the <3.5 g/dL threshold across both raw ($p = 0.0159$) and KNN-imputed datasets ($p = 0.0110$). In the raw cohort, there were no benign cases with lower levels of albumin, which led to an infinite point estimate (O.R. = infinite) and uninformative confidence bounds. While out-of-fold KNN imputation stabilized the distribution to produce a finite estimate (O.R. = 37.67), the resulting 95% confidence interval was still quite wide [2.98, 476.20] (Figures 5 & 6). This interval expansion is caused by Wald-type log-odds approximations under low cell frequency regimes, where small event counts can increase the standard error matrix. Therefore, we can conclude that hypoalbuminemia (<3.5 g/dL) increases the odds of lung cancer by at least three times (O.R. > 2.98). The extremely wide upper limit is caused by the small number of patients with low albumin rather than representing a true 400-fold risk increase. Clinically, this association may exist because, in cancerous cases, systemic/chronic inflammation mediated by tumors often limits albumin production. With the demanding metabolic activities of cancer, serum albumin may be further depleted. However, it is important to note that numerous non-cancer-related variables, including chronic disease, hydration status, and overall nutritional health, can influence albumin levels. As a result, albumin may function better as a prognostic indicator in patients with established malignancy rather than as a diagnostic predictor in early-stage/screening contexts. Additionally, the relatively small number of malignant cases in the dataset may have limited the statistical power required to detect a meaningful association.

Logistic Regression & ROC Curve Analysis

The logistic regression model further explored the predictive potential of these markers as a unified risk assessment tool. By assigning weighted coefficients to each variable, the model quantified how each marker contributes to the overall probability of malignancy. The accuracy of this binary classification model was assessed using a receiver operating characteristic (ROC) curve and its area under the curve (AUC). The AUC of the ROC curve was 0.53, indicating an accuracy only slightly higher than random guessing (0.5 = no-skill model).

The optimal threshold was 0.19, meaning that the probability cutoff was used to classify a case as malignant or benign. Values above this point are labeled malignant, while values below are considered benign. This threshold maximizes accuracy in identifying positive and negative cases in this dataset. At this threshold, the model achieved a Youden's index of 0.3126, in which higher values indicate better classification ability, while 0 indicates poor performance. The model demonstrated high specificity (0.939), meaning it can correctly identify benign cases 93.9% of the time, but low sensitivity (0.375), meaning it can correctly identify 37.5% of malignant cases. However, given the AUC (0.53), these sensitivity/specificity values are not evidence of diagnostic utility. The low sensitivity means a benign prediction from this model cannot be relied upon to exclude malignancy, and the high specificity is more likely a result of the uneven amounts of cancerous versus benign training data. There were only 8

malignant cases for the model to work from, while 114 benign cases were available to guide the machine learning model. This lack of information on malignant cases suggests that with more cancerous cases for the program to work from, a more precise predictive model could be developed.

This current model should be viewed as exploratory rather than definitive, as there is just not enough data to be able to surmise definite conclusions about the involvement of clinical markers as a pure way to evaluate a patient's risk of developing lung cancer.

CONCLUSION

While the p-values of some individual clinical markers did not demonstrate statistically significant associations with lung cancer, serum albumin (<3.5 g/dL) did show a significant association, as did WBC count post-KNN adjustment. To evaluate whether combining these features could improve performance, a multivariate approach was tested. However, combining these markers into a logistic regression model yielded limited discriminative ability (AUC = 0.53). The high specificity (0.939) but low sensitivity (0.375) of the model suggests that the model does not have meaningful clinical utility, alongside the near-chance AUC, which discounts any diagnostic value the specificity and sensitivity figures might otherwise seem to suggest. Overall, the findings support the conclusion that use of SpO₂, WBC count, and serum albumin combined are insufficient as diagnostic tools. Larger datasets and refined modeling may contribute to a more accurate binary classification of malignant versus benign cases in LR0 patients in the future.

Limitations

Despite the associations found between some clinical markers and malignancy, there are limitations that must be addressed. One of these limitations includes selection bias in the dataset. In a real-world healthcare setting, patients typically obtain clinical marker data only when an underlying issue requires them to seek care in the first place. Due to this variable, there are very few “random” cases in which the patient is completely fine without any other health conditions that may be causing their visit. Even if a patient was getting their routine check-up, most clinical markers, such as WBC count, SpO₂, and Serum Albumin, are not being measured. This factor may skew the results because elevated or depressed nonspecific clinical markers can be associated with other morbidities. Similarly, many patients getting an LDCT LCS either used to be, or are current smokers, and are over 50 years old. These extrinsic factors could also skew the significance of each of the markers measured to be greater; smoking is widely known as a common carcinogen that leads to cancer, especially lung cancer, and old age is linked to a higher risk of cancer as the body's DNA repair mechanisms get weaker with time.

Selection bias is also introduced into the study due to the need for patients to be followed up. The patients who did come back for follow-up probably only did so because they had continuing symptoms, the doctors were worried about them, or there may have been some other reason that is linked to having cancer. Patients that were not followed up (transferred out of IU Health, etc.), no matter if they had cancer or not, were removed from the study. Therefore, the number of cancers found in these patients may not

represent the actual number of lung cancers present in patients classified as LR0 and can either overestimate or underestimate the actual rate of cancer in these patients based on why those patients that were not followed up were removed from the study.

A large limitation that persisted throughout the duration of this study was the small sample size and the lack of representativeness of populations due to the scarce amount of data. There were 122 cases available for potential use. From these potential cases, 65 had available SpO₂ values, 30 had WBC count values, and 28 had available serum albumin values. This low number of cases reduces the generalizability of the results and creates a less accurate diagnostic model.

Implications

Based on the results of this study, physicians should avoid using nonspecific clinical markers in their diagnosis due to their lack of precision. There are thousands of morbidities associated with elevated/depressed SpO₂, WBC count, and serum albumin. Relying on such broad indicators introduces clinical ambiguity, making it difficult to extrapolate lung cancer from simple clinical markers. By reallocating research funding and efforts toward targeted radiological models and specific biomarkers, the medical community can move past "diagnostic noise." Currently, there are efforts into finding lung-cancer specific biomarkers, in which the PD-L1 pathway has been found in over 50% of cancer cells in non-small cell lung cancer (Han et al., 2020). Showing promise that other biomarkers are available to aid in lung cancer diagnosis.

Future Research

Future research should focus on enhancing the model's predictive power by incorporating a larger and more diverse dataset. Rather than serving as a standalone tool, future studies should focus on integrating variables such as smoking history, nodule size, and specific radiological imaging features with multi-institutional and more robust datasets to gain a more comprehensive representation of patient risk. Currently, the most accurate model for lung cancer prediction is the Herder model with an ROC-AUC of 0.83 (Susam et al., 2022). Furthermore, these performances can be compared through prospective studies to determine if a hybrid modeling approach can surpass current diagnostic standards by combining existing models. Ultimately, this transition from retrospective analysis to prospective clinical validation will assess the model's efficacy in predicting lung cancer.

Conclusion

This study highlights the complexity of improving diagnostic accuracy in LR0 cases and the limitations of relying on nonspecific clinical markers. While the current findings suggest limited predictive ability, they open the door for more refined approaches that integrate clinical, radiological, and molecular data. Moving forward, expanding datasets and validating models will be essential to strengthening predictive performance. Continued research in this area is strongly encouraged, particularly in exploring specific biomarkers and hybrid MPMs. Ultimately, collaboration between researchers and physicians will be critical in advancing more precise tools, and further discussion is welcome to build upon these findings and drive innovation in lung cancer diagnostics.

REFERENCES

- Akcam, T.I., Tekneci, A.K., Turhan, K. et al. Prognostic value of systemic inflammation markers in early-stage non-small cell lung cancer. *Sci Rep* 15, 33886 (2025).
<https://doi.org/10.1038/s41598-025-08683-y>
- Arora, S., Hammer, M. M., & Byrne, S. C. (2025). Real-World Use and Outcomes of Lung-RADS v1.1 Category 4B Versus Lung-RADS v2022 Category 0 for Suspected Infectious or Inflammatory Findings on Lung Cancer Screening CT. *AJR. American journal of roentgenology*, 225(5), e2533279. <https://doi.org/10.2214/AJR.25.33279>
- Azour, L., Moore, W. H., O'Donnell, T., Truong, M. T., Babb, J., Niu, B., Wimmer, A., Kiumehr, S., & Ko, J. P. (2022). Inter-reader variability of volumetric Subsolid pulmonary nodule Radiomic features. *Academic Radiology*, 29, S98-S107. <https://doi.org/10.1016/j.acra.2021.01.026>
- Cancer of the lung and bronchus - Cancer Stat facts. (2026). SEER.
https://seer.cancer.gov/statfacts/html/lungb.html#xd_co_f=YTM0NDM1ZTgtZmQzMj00ZDY4LTliMjEtMmFiYzkyNWZjZmE3~
- Chen Y, Liu T, Feng H, Liu T, Zhang J, Wang J, Lu J, Rossi A, Riano I, Hu P, Zhang J. The prognostic role of albumin levels in lung cancer patients receiving third-line or advanced immunotherapy: a retrospective study. *Transl Lung Cancer Res*. 2024 Jun 30;13(6):1307-1317. doi: 10.21037/tlcr-24-378. Epub 2024 Jun 21. PMID: 38973954; PMCID: PMC11225039.
- Chudgar NP, Bucciarelli PR, Jeffries EM, Rizk NP, Park BJ, Adusumilli PS, Jones DR. Results of the national lung cancer screening trial: where are we now? *Thorac Surg Clin*. 2015 May;25(2):145-53. doi: 10.1016/j.thorsurg.2014.11.002. Epub 2015 Feb 2. Erratum in: *Thorac Surg Clin*. 2020 May;30(2):xi. doi: 10.1016/j.thorsurg.2020.02.002. PMID: 25901558; PMCID: PMC4817217.
- Common cancer sites - Cancer Stat facts. (2026). SEER.
<https://seer.cancer.gov/statfacts/html/common.html>
- Han Y, Liu D, Li L. PD-1/PD-L1 pathway: current researches in cancer. *Am J Cancer Res*. 2020 Mar 1;10(3):727-742. PMID: 32266087; PMCID: PMC7136921.
- Hu, Z., Wu, W., Zhang, X., Li, P., Zhang, H., Wang, H., Xue, W., Chen, Z., Zhao, Q., & Duan, G. (2021). Advanced Lung Cancer Inflammation Index is a Prognostic Factor of Patients with Small-Cell Lung Cancer Following Surgical Resection. *Cancer management and research*, 13, 2047–2055.
<https://doi.org/10.2147/CMAR.S295952>
- Isaksson, J., Wennström, L., Branden, E., Koyi, H., Berglund, A., Micke, P., Mattsson, J. S., Willén, L., & Botling, J. (2022). Highly elevated systemic inflammation is a strong independent predictor of early mortality in advanced non-small cell lung cancer. *Cancer Treatment and Research Communications*, 31, 100556. <https://doi.org/10.1016/j.ctarc.2022.100556>
- Li, J., Guo, S., Ma, R. et al. Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. *BMC Med Res Methodol* 24, 41 (2024).
<https://doi.org/10.1186/s12874-024-02173-x>

- Martins SJ, Ho N, Cavamura SO, Harada CM, Yamamoto CA, Takagaki TY. Lung cancer symptoms and pulse oximetry in the prognostic assessment of patients with lung cancer. *BMC Cancer*. 2005 Jul 6;5:72. doi: 10.1186/1471-2407-5-72. PMID: 16000170; PMCID: PMC1177931.
- Mendoza, D. P., Chintanapakdee, W., Zhang, E. W., Gilman, M. D., Lennes, I. T., Frank, A. J., Shepard, J. O., & Digumarthy, S. R. (2021). Management and Outcomes of Suspected Infectious and Inflammatory Lung Abnormalities Identified on Lung Cancer Screening CT. *AJR. American journal of roentgenology*, 217(5), 1083–1092. <https://doi.org/10.2214/AJR.20.25124>
- McWilliams, A., Tammemagi, M. C., Mayo, J. R., Roberts, H., Liu, G., Soghrati, K., Yasufuku, K., Martel, S., Laberge, F., Gingras, M., Atkar-Khattra, S., Berg, C. D., Evans, K., Finley, R., Yee, J., English, J., Nasute, P., Goffin, J., Puksa, S., Stewart, L., ... Lam, S. (2013). Probability of cancer in pulmonary nodules detected on first screening CT. *The New England journal of medicine*, 369(10), 910–919. <https://doi.org/10.1056/NEJMoa1214726>
- Pavlou, A., Kurtz, R. M., & Song, J. W. (2021). Diagnostic Accuracy Studies in Radiology: How to Recognize and Address Potential Sources of Bias. *Radiology research and practice*, 2021, 5801662. <https://doi.org/10.1155/2021/5801662>
- Şahin, F., & Aslan, A. F. (2018). Relationship between Inflammatory and Biological Markers and Lung Cancer. *Journal of clinical medicine*, 7(7), 160. <https://doi.org/10.3390/jcm7070160>
- Susam, S., Çinkooğlu, A., Ceylan, K. C., Gürsoy, S., Kömürçüoğlu, B. E., Mertoğlu, A., Çırak, A. K., Gayaf, M., Güldaval, F., Tuksavul, F., Polat, G., Ataman, S., Yıldırım, E., Koparal, H., & Yücel, N. (2022). Comparison of Brock University, Mayo Clinic and Herder models for pretest probability of cancer in solid pulmonary nodules. *The clinical respiratory journal*, 16(11), 740–749. <https://doi.org/10.1111/crj.13546>
- Winther-Larsen A, Aggerholm-Pedersen N, Sandfeld-Paulsen B. Inflammation-scores as prognostic markers of overall survival in lung cancer: a register-based study of 6,210 Danish lung cancer patients. *BMC Cancer*. 2022 Jan 14;22(1):63. doi: 10.1186/s12885-021-09108-5. PMID: 35027001; PMCID: PMC8759208.
- Wong JYY, Bassig BA, Loftfield E, Hu W, Freedman ND, Ji BT, Elliott P, Silverman DT, Chanock SJ, Rothman N, Lan Q. White Blood Cell Count and Risk of Incident Lung Cancer in the UK Biobank. *JNCI Cancer Spectr*. 2019 Dec 12;4(2):pkz102. doi: 10.1093/jncics/pkz102. PMID: 33313477; PMCID: PMC7083262.
- Yang, Z., Zheng, Y., Wu, Z., Wen, Y., Wang, G., Chen, S., Tan, F., Li, J., Wu, S., Dai, M., Li, N., & He, J. (2021). Association between pre-diagnostic serum albumin and cancer risk: Results from a prospective population-based study. *Cancer medicine*, 10(12), 4054–4065. <https://doi.org/10.1002/cam4.3937>
- Zhai Y, Wu J, Tang C, Huang B, Bi Q, Luo S. Characterization of blood inflammatory markers in patients with non-small cell lung cancer. *Int J Clin Exp Pathol*. 2024 May 15;17(5):165-172. doi: 10.62347/IPTW9741. PMID: 38859920; PMCID: PMC11162609.
- Zhang Y, Liu W. The CALLY index in cancer research: a comprehensive review. *Oncol Rev*. 2026 Feb 19;20:1737805. doi: 10.3389/or.2026.1737805. PMID: 41798270; PMCID: PMC12960488.
- Zhou T, He X, Fang W, Zhan J, Hong S, Qin T, Ma Y, Sheng J, Zhou N, Zhao Y, Huang Y, Zhang L. Pretreatment Albumin/Globulin Ratio Predicts the Prognosis for Small-Cell Lung Cancer.

Medicine (Baltimore). 2016 Mar;95(12):e3097. doi: 10.1097/MD.0000000000003097. PMID: 27015181; PMCID: PMC4998376.