

Paper Type: Original Article

Q-LungBench: Leakage-Aware Benchmarking of Machine Learning Models for Questionnaire-Based Lung Cancer Risk Classification

Mohamed Eassa¹  and Ahmed Abdelhafeez^{2,*} ¹ Department of Computer Science, Faculty of Information Systems and Computer Science, October 6 University, Giza 12585, Egypt; Mohamed.eassa.cs@o6u.edu.eg.² Faculty of Computers and Information Technology, Innovation University, Cairo, Egypt; Ahmed.abdelhafeez@iu.edu.eg.**Received:** 02 Jan 2026**Revised:** 25 Apr 2026**Accepted:** 28 Jul 2026**Published:** 30 Jul 2026**Abstract**

Questionnaire-based lung cancer datasets are widely used in machine-learning studies, yet their small size, class imbalance, duplicate records, and test-guided model choices can make performance appear better than it is. We developed Q-LungBench as a reproducible, leakage-aware framework for evaluating six classifiers on the public Survey Lung Cancer dataset. After exact duplicates were removed before data partitioning, 276 unique records with 15 predictors remained. The data were divided using a stratified 80:20 development-test split. On the development set only, logistic regression, radial-basis-function support vector machine, random forest, gradient boosting, k-nearest neighbours, and Gaussian naive Bayes were tuned with five-fold stratified cross-validation. The prespecified selection criterion, mean balanced accuracy, identified logistic regression as the final model (0.854) before independent testing. On the 56-record test set, logistic regression achieved 91.07% accuracy, 94.79% balanced accuracy, 89.58% sensitivity, 100% specificity, an F1-score of 94.51%, a Matthews correlation coefficient of 0.742, a ROC-AUC of 0.971 (95% bootstrap confidence interval 0.917-1.000), and a PR-AUC of 0.995. Permutation analysis highlighted allergy, swallowing difficulty, wheezing, coughing, and chest pain as the variables on which the fitted model relied most. The findings indicate that transparency in the assessment process does not necessarily mean inferiority without any fusion or artificial oversampling afterward. However, the limited size of negative instances and the uncertain source of the dataset prevent us from making any diagnosis at this point.

Keywords: Lung Cancer; Machine Learning; Questionnaire Data; Data Leakage; Reproducibility; Model Evaluation.

1 | Introduction

Despite great advancements in medicine, lung cancer still represents the most frequently occurring type of cancer and is the major cause of cancer-associated mortality in the world. According to GLOBOCAN, there were about 2.5 million cases and 1.8 million deaths from lung cancer in 2022, and the recent reports of the WHO still stress the problems of late diagnostics and the necessity for risk-based screening [1, 2]. Diagnostics is conducted according to the clinical pathway validated by medical institutions and can include such methods as low-dose computed tomography, imaging, biopsy, molecular testing, and the opinion of various specialists. A questionnaire-based machine learning model can never substitute any of these processes.

**Corresponding Author:** Ahmed.abdelhafeez@iu.edu.eg<https://doi.org/10.61356/j.scin.2026.3665>Licensee **SciNexuses**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

The Public Lung Cancer survey data has been widely used to compare the performance of traditional machine learning and deep learning approaches. In this regard, Wani et al. proposed DeepXplainer, which is based on a CNN-XGBoost model with SHAP explanations, achieving an accuracy of 97.43% [3]. Maurya et al. analyzed a few algorithms with different techniques of class balancing [4]. Pavithran et al. considered various combinations of augmentation methods along with classifier, where the highest accuracy of 93.55% and 96.76% ROC-AUC was achieved using K-Means SMOTE and Multilayer Perceptron [5]. Ayad et al. considered Recursive Feature Elimination along with Support Vector Machines and XGBoost optimization in two datasets [6]. Despite their effectiveness, none of the aforementioned works are comparable. The handling of duplicates, data splitting, resampling, model selection, and repeated use of the test set differs from one pipeline to another.

These concerns become particularly important when the dataset is small. If duplicate records are divided between training and testing sets, or if preprocessing and oversampling are completed before cross-validation, information can leak into the evaluation process. A similar problem arises when researchers inspect test results and then use them to choose or refine the final model. In each case, the reported performance may be overly optimistic. The TRIPOD+AI statement therefore calls for clear reporting of the data source, model-development process, performance measures, and open-science materials [7].

Rather than proposing another complex classifier, Q-LungBench focuses on the integrity of the evaluation process. We make five contributions. First, exact duplicates are removed before the data are partitioned. Second, model development is kept separate from independent testing. Third, hyperparameters are optimized by cross-validation using balanced accuracy as the prespecified selection criterion. Fourth, performance is described with discrimination measures, class-sensitive metrics, probability error, bootstrap confidence intervals, and a paired model comparison. Finally, the code, figures, summary tables, and run metadata are released publicly. Our central question is straightforward: can a transparent conventional model remain competitive without post-hoc fusion or synthetic data generation?

2 | Related Work

Artificial-intelligence research in lung cancer now includes radiological images, electronic health records, molecular measurements, and questionnaire data. Imaging models can learn complex visual patterns, but they usually require large, carefully curated datasets and fall outside the scope of this work. Results from structured questionnaires are more easily analyzed and can serve for preliminary studies in methodology. Yet their actual performance is highly dependent on both the source of the cohort and design of the study.

Table 1. Positioning of Q-LungBench in methodology compared to recent studies using structured data.

Study	Approach	Reported result	Primary emphasis
Wani et al. [3]	Survey Lung Cancer; CNN-XGBoost with SHAP	Accuracy 97.43%; sensitivity 98.71%; F1 98.08%	Focus on architecture and explainability
Maurya et al. [4]	Comparative ML with class balancing	Multiple classifiers compared	Algorithm comparison with class balancing
Pavithran et al. [5]	Nine resampling methods with multiple classifiers and LIME	Accuracy 93.55%; ROC-AUC 96.76%	Interaction between augmentation and classifier choice
Ayad et al. [6]	RFE-SVM plus Nelder-Mead optimized XGBoost	Reported 100% accuracy across two datasets	Feature selection and optimization
Q-LungBench	Six optimized conventional models; no synthetic oversampling	Independent-test ROC-AUC 0.971; balanced accuracy 0.948	Leakage control and reproducibility

Q-LungBench is included in Table 1 for the purposes of comparing our study against more recent research in terms of methodology rather than ranking. Due to differences in random splitting, duplicates removal,

sampling, validation and class imbalance in datasets, quantitative comparison would be misleading. Every result is considered dataset-specific.

3 | Materials and Methods

3.1 | Dataset and Outcome

The data source used was the publicly available Survey Lung Cancer CSV file from an open data mirror [8]. Prior to cleaning, the data set had 309 observations, 15 predictors, and one binary dependent variable known as LUNG CANCER. The predictors included gender, age, smoker, yellow fingers, anxiety, peer pressure, chronic disease, fatigue, allergy, wheezing, alcohol consuming, coughing, shortness of breath, swallowing, and chest pain. We encoded the positive class as 1 and the negative class as 0. Because the dataset contains questionnaire responses rather than verified imaging, pathology, stage, or longitudinal outcomes, we treat the outcome as a benchmark label rather than a clinically confirmed diagnosis.

Table 2. Dataset and partition profile.

Characteristic	Value
Raw records	309
Exact duplicates removed	33
Unique records analysed	276
Positive class	238 (86.23%)
Negative class	38 (13.77%)
Predictors	15
Development set	220 records
Independent test set	56 records (48 positive, 8 negative)

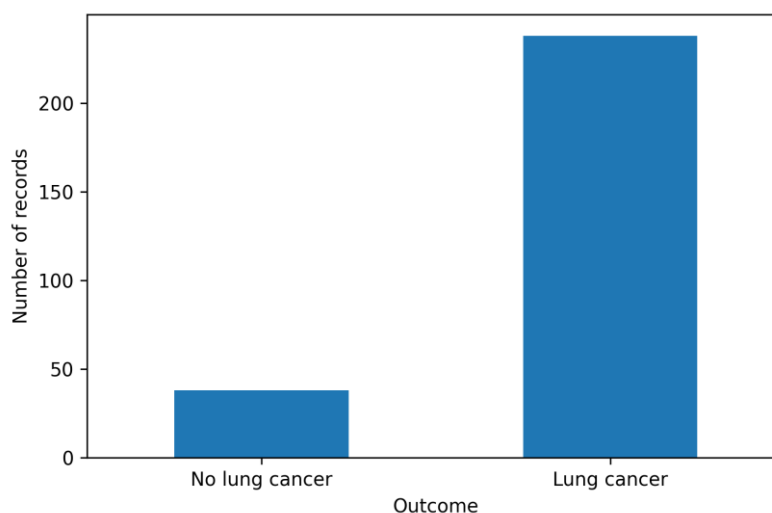


Figure 1. Class distribution after exact duplicate records were removed.

3.2 | Preprocessing and Leakage Control

We first standardized the column names and converted the target to binary form. Gender was encoded as male=1 and female=0, while questionnaire items originally coded as 1 and 2 were recoded as 0 and 1. Exact duplicate rows were then removed before any data split. Using random state 42, we divided the remaining records by stratified sampling into an 80% development set and a 20% independent test set. The test data were kept out of hyperparameter tuning and model ranking. Within each modelling pipeline, age was median-imputed and standardized, whereas the binary variables were imputed with their most frequent value. No

values were missing in the analysed file, so the imputation steps served only to make the workflow robust and reproducible.

We did not apply synthetic oversampling. Logistic regression, RBF-SVM, and random forest were fitted with balanced class weights; the other models used their stated native configurations. This choice avoided creating synthetic observations from a small dataset whose original clinical provenance is not fully documented.

3.3 | Candidate Models and Optimization

We compared six classifiers: logistic regression, an RBF support vector machine, random forest, gradient boosting, k-nearest neighbours, and Gaussian naive Bayes. Together, they represent linear, kernel-based, bagging, boosting, instance-based, and probabilistic approaches [9-13]. Hyperparameters were tuned through grid search with five-fold stratified cross-validation on the development set. Mean balanced accuracy was chosen in advance as the refit and model-selection criterion because ordinary accuracy can hide poor performance in the smaller negative class. We also recorded ROC-AUC, PR-AUC, and F1-score during cross-validation. Once selected, the final model was fixed before the independent test set was evaluated.

Table 3. Five-fold cross-validation model-selection results on the development set.

Model	Bal. acc.	ROC-AUC	F1	PR-AUC
LR	0.854 ± 0.073	0.910	0.920	0.985
SVM	0.843 ± 0.067	0.924	0.907	0.988
RF	0.779 ± 0.083	0.921	0.948	0.987
GB	0.765 ± 0.083	0.826	0.948	0.958
KNN	0.735 ± 0.046	0.771	0.931	0.938
GNB	0.721 ± 0.107	0.874	0.932	0.978

The selected settings were: logistic regression, $C=0.01$; RBF-SVM, $C=1$ and $\gamma=0.01$; random forest, unlimited depth, a minimum leaf size of 2, and 400 trees; gradient boosting, a learning rate of 0.1, a maximum depth of 2, and 200 trees; k-nearest neighbours, $k=3$ with distance weighting; and Gaussian naive Bayes, variance smoothing of $1e-11$.

3.4 | Evaluation and Statistical Analysis

Performance on the independent test set was assessed with accuracy, balanced accuracy, sensitivity, specificity, precision, F1-score, Matthews correlation coefficient, ROC-AUC, PR-AUC, and Brier score. We estimated 95% confidence intervals for ROC-AUC with 1,000 nonparametric bootstrap resamples [14-17]. ROC and precision-recall curves were produced for all six models. To compare the two highest-ranked models from cross-validation on the same test cases, we used the exact McNemar test [18]. We also calculated model-agnostic permutation importance for the selected model by measuring the loss in balanced accuracy after 100 permutations of each feature. These are the importance scores of how dependent the model is on that particular variable for this sample; they do not imply any causality.

3.5 | Reproducibility and Open Science

The analysis was coded using Python with scikit-learn [19]. Q-LungBench provides the source code, a Google Colab notebook, dependency files, tables and figures, run information, and citations [20]. In order to prevent re-distribution of the record level data, the repository accesses the public CSV upon execution of the code or asks the user to provide it manually. A fixed random state, saved split indices in the experiment archive, and tables of results allow for reproduction of the reported analysis.

4 | Results

4.1 | Cross-validation Model Selection

For the purpose of selection prior to analyzing the test data set, logistic regression yielded the highest mean balanced accuracy through cross-validation (0.854 ± 0.073). Next in terms of balanced accuracy was RBF-SVM (0.843 ± 0.067) while also achieving the highest ROC-AUC (0.924) during cross-validation. The random forest algorithm and gradient boosting yielded high F1 scores but low balanced accuracy, which means that they were better at dealing with the predominant class.

4.2 | Independent-test Performance

Table 4. Classification performance on the independent test set (all values are proportions).

Model	Acc.	Bal. Acc.	Sens.	Spec.	F1
LR	0.911	0.948	0.896	1.000	0.945
SVM	0.875	0.927	0.854	1.000	0.921
RF	0.929	0.854	0.958	0.750	0.958
KNN	0.929	0.854	0.958	0.750	0.958
GNB	0.893	0.729	0.958	0.500	0.939
GB	0.875	0.719	0.938	0.500	0.928

Table 5. Discrimination and probability error on the independent test set.

Model	MCC	ROC-AUC (95% CI)	PR-AUC	Brier
LR	0.742	0.971 (0.917-1.000)	0.995	0.191
SVM	0.675	0.977 (0.932-1.000)	0.996	0.051
RF	0.708	0.966 (0.917-1.000)	0.995	0.059
KNN	0.708	0.898 (0.743-0.998)	0.972	0.069
GNB	0.519	0.953 (0.889-0.997)	0.993	0.075
GB	0.463	0.906 (0.793-0.987)	0.984	0.089

The selected logistic-regression model correctly classified 51 of the 56 test records. It made no false-positive predictions among the eight negative cases and correctly identified 43 of the 48 positive cases, leaving five false negatives. This corresponded to 91.07% accuracy, 94.79% balanced accuracy, 89.58% sensitivity, 100% specificity, 100% precision, an F1-score of 94.51%, and an MCC of 0.742. The ROC-AUC was 0.971 (95% bootstrap CI 0.917-1.000), and the PR-AUC was 0.995. The apparently perfect specificity should be read cautiously, because it is based on only eight negative test records.

Random forest and k-nearest neighbours both reached a slightly higher raw accuracy of 92.86%. However, each model misclassified two of the eight negative cases, which reduced specificity to 75% and balanced accuracy to 85.42%. RBF-SVM achieved the highest ROC-AUC (0.977) and the lowest Brier score (0.051), although its sensitivity was lower than that of logistic regression. This pattern shows why the model-selection rule should be defined on the development data rather than choosing whichever model happens to have the most attractive test-set metric.

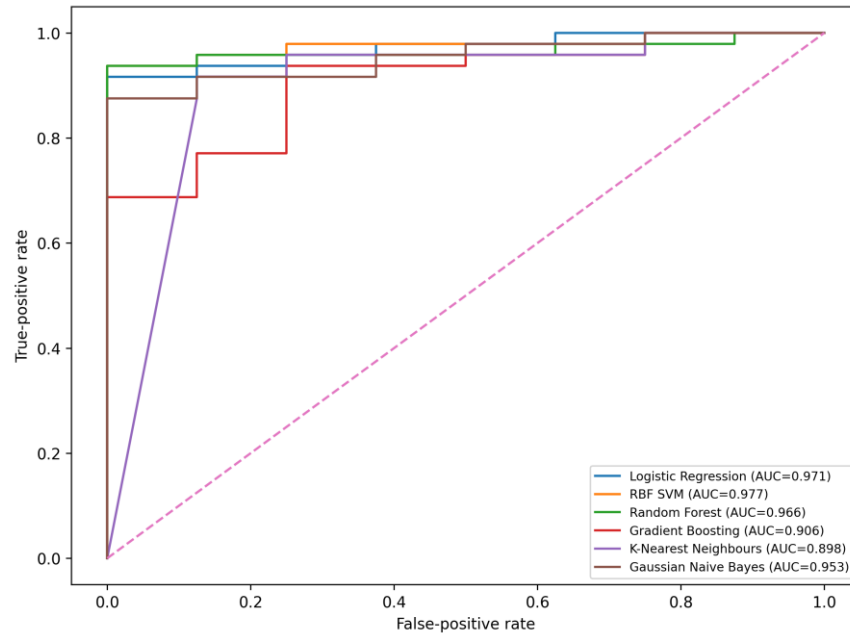


Figure 2. Receiver operating characteristic curves for the independent test set.

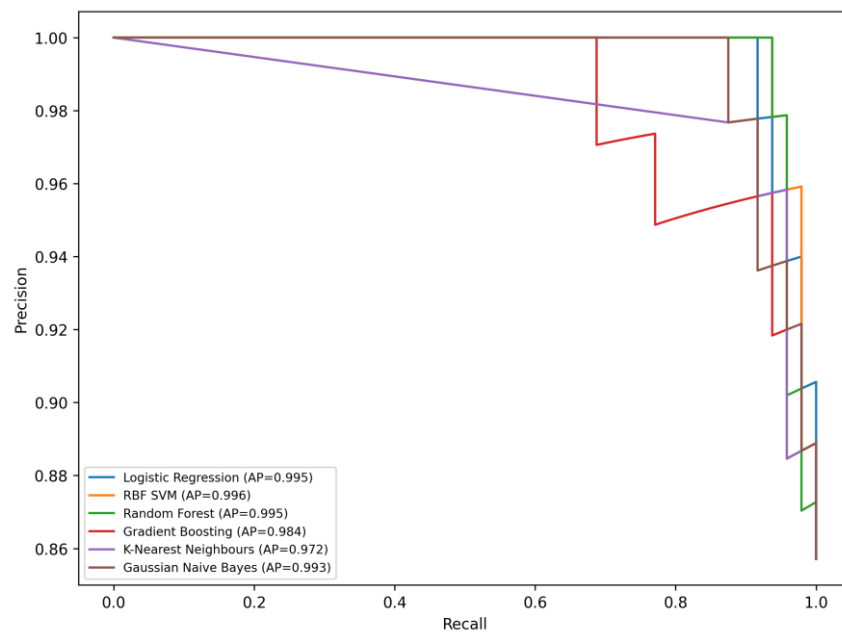


Figure 3. Precision-recall curves for the independent test set.

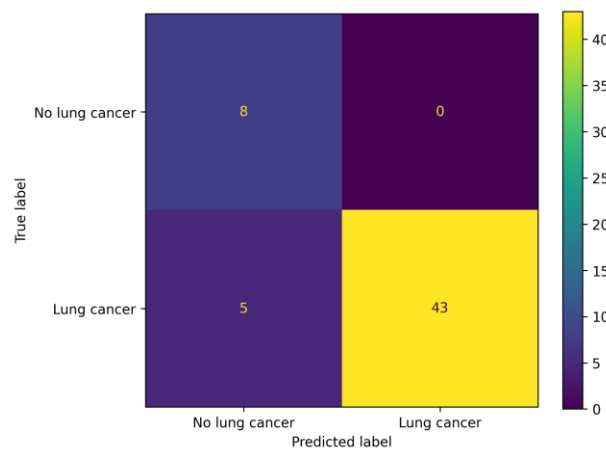


Figure 4. Confusion matrix for the logistic-regression model selected by cross-validation.

4.3 | Statistical Comparison and Variable Importance

In the McNemar Exact Test, logistic regression was able to correctly identify only two cases, while RBF-SVM failed to correctly identify them; conversely, RBF-SVM was unable to correctly identify any case that logistic regression could not. This resulted in a p-value of 0.500, and with such few numbers, no statistically reliable differences exist. Permutation importance identified allergy (0.102), swallowing difficulty (0.091), wheezing (0.080), coughing (0.078), chest pain (0.071), yellow fingers (0.060), alcohol consumption (0.051), and fatigue (0.040) as the variables associated with the largest mean decreases in balanced accuracy. These estimates came from only 56 test cases, and several questionnaire variables may be correlated. The ranking should therefore be understood as a description of this model's behaviour in this sample, not as evidence that these variables cause lung cancer.

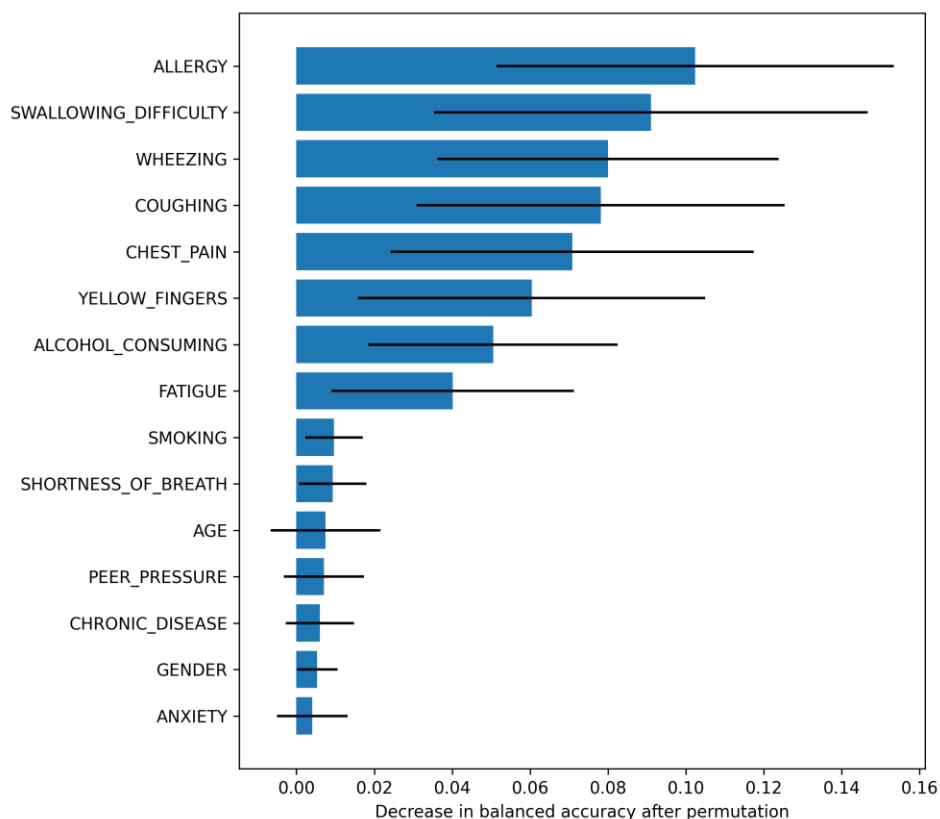


Figure 5. Permutation importance for the selected logistic-regression pipeline.

5 | Discussion

Q-LungBench shows that a conventional regularized logistic-regression model can discriminate questionnaire labels well when model development and testing are kept strictly separate. Without synthetic oversampling, deep neural networks, feature fusion, or post-hoc model selection, the chosen model reached a test ROC-AUC of 0.971 and a balanced accuracy of 0.948. However, one cannot consider this as evidence of the overall superiority of the logistic regression. This is just an example of how a well-defined baseline can hold up against a small and structured dataset.

It is worth noting that the performance is consistent with other recent research works that relied on the Survey Lung Cancer dataset. According to DeepXplainer, the accuracy of CNN-XGBoost architecture was 97.43% [3]. Pavithran et al. managed to reach 93.55% accuracy and 96.76% AUC with the help of K-Means SMOTE + MLP [5]. Q-LungBench showed somewhat worse raw accuracy compared to DeepXplainer but comparable AUC to the augmentation-based baseline. Yet, the comparison of their numerical values would be meaningless due to the fact that different datasets were considered by each of them. Thus, the value of Q-LungBench does not lie in state-of-the-art accuracy.

However, the distinction between ordinary accuracy and balanced accuracy was not insignificant in this case. Although random forest and k-nearest neighbours were the two most accurate models on test data, they also produced the greatest number of errors in the negative class. Either of these models would be considered the winner after seeing the results from test data. However, the winner chosen based on the balanced accuracy calculated on the development data was the logistic regression, which stayed the winner after being evaluated on test data. Moreover, RBF-SVM also achieved slightly better ROC-AUC value on test data as well as significantly lower Brier score, but only after the model selection process was over.

Allergy, swallowing difficulty, wheezing, coughing, chest pain, yellow fingers, alcohol consumption, and fatigue received the largest permutation-importance scores. Several of these variables are plausible symptoms or exposure-related correlates, but an association in a questionnaire dataset is not the same as etiological importance. Smoking, for example, received a relatively small score. This is more likely to reflect the composition of the sample, correlation between predictors, the coding scheme, and the benchmark label than any challenge to the established causal relationship between smoking and lung cancer. Familiar clinical terms do not, by themselves, turn a convenience dataset into a validated medical instrument.

We deliberately did not add fusion or more aggressive optimization after seeing the independent-test results. Building an ensemble in response to those results would allow the test data to influence model development and would no longer leave them as an untouched evaluation set. Any future ensemble, calibration method, or feature-selection procedure should be specified before a new external cohort is examined. Alternatively, it could be assessed through repeated nested cross-validation.

5.1 | Limitations

Some important limitations in the current study include the sample size; after the removal of duplicates, there were only 276 observations left, among which 38 were labeled as negatives, and the test set included only eight negatives. Therefore, the confidence intervals become large, and the paired statistical tests become underpowered. Moreover, there is very poor documentation in the dataset with regard to recruitment, case confirmation, the setting of the study, missing data mechanisms, and the eligibility criteria. Apart from exact duplicates, we cannot exclude the possibility that some of the observations may come from the same patients. Another limitation includes the higher prevalence of positive labels compared to what would normally be observed in a general screening population. There is no other dataset to assess any of the factors.

5.2 | Practical Implications

Q-LungBench is meant to be an educational tool and a benchmarking exercise. It serves as a useful case study that could be used to illustrate concepts of data leakage, class imbalance, model selection, and reproducible reporting. Q-LungBench should never be used to provide comfort, diagnosis, screening, or triage for any individual patient. For a clinically relevant follow-up study, there should be prospective eligibility criteria, confirmed outcome, representative enrollment, external validation, formal calibration, decision-curve analysis, and comparison to other screening methods.

6 | Conclusion

Q-LungBench provides a reproducible, leakage-aware comparison of six machine-learning models for questionnaire-based lung cancer risk classification. After exact duplicate removal and cross-validated model selection, logistic regression achieved 94.79% balanced accuracy and an ROC-AUC of 0.971 on the independent test set. The study illustrates that careful separation of development and evaluation, together with complete reporting, can be more informative than adding complexity after the results are known. Although the performance is useful as a methodological benchmark, it is not sufficient to support clinical use. The next essential step is external validation in large, representative cohorts with diagnostically verified outcomes.

Acknowledgments

We thank the maintainers of the public Survey Lung Cancer dataset and the contributors to the open-source Python scientific-computing ecosystem.

Author Contribution

Conceptualization: M.E. and A.A.; methodology: M.E. and A.A.; software: M.E.; validation: A.A.; formal analysis: M.E.; investigation: M.E.; data curation: M.E.; visualization: M.E.; supervision and project administration: A.A.; original draft preparation: M.E.; review and editing: A.A. Both authors read and approved the submitted manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Ethics committee approval was not required because this study analysed a publicly available, secondary, de-identified dataset and did not involve direct contact with human participants.

Informed Consent Statement

Not applicable. This study was a secondary analysis of a public, de-identified dataset.

Data Availability

The analysis was based on the public Survey Lung Cancer dataset [8]. Reproducibility code, the Google Colab notebook, summary tables, figures, run metadata, and dependency specifications are available through the Q-LungBench GitHub repository [20]. Raw and cleaned record-level data are not redistributed. Instead, the code retrieves the public CSV at runtime or asks the user to upload it manually.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] World Health Organization. Lung cancer. WHO Fact Sheet. Updated 16 April 2026.
- [2] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74, 229-263. <https://doi.org/10.3322/caac.21834>
- [3] Wani, N. A., Kumar, R., & Bedi, J. (2024). DeepXplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence. *Computer Methods and Programs in Biomedicine*, 243, 107879. <https://doi.org/10.1016/j.cmpb.2023.107879>
- [4] Maurya, S. P., Sisodia, P. S., Mishra, R., & Singh, D. P. (2024). Performance of machine learning algorithms for lung cancer prediction: A comparative approach. *Scientific Reports*, 14, 18562. <https://doi.org/10.1038/s41598-024-58345-8>
- [5] Pavithran M. S., Saranyaraj, D., & Chakraborty, A. (2025). Lung cancer risk prediction using augmented machine learning pipelines with explainable AI. *Frontiers in Artificial Intelligence*, 8, 1602775. <https://doi.org/10.3389/frai.2025.1602775>
- [6] Ayad, S., Al-Jamimi, H. A., & El Kheir, A. (2025). Integrating advanced techniques: RFE-SVM feature engineering and Nelder-Mead optimized XGBoost for accurate lung cancer prediction. *IEEE Access*, 13, 29589-29600. <https://doi.org/10.1109/ACCESS.2025.3536034>
- [7] Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., et al. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385, e078378. <https://doi.org/10.1136/bmj-2023-078378>
- [8] Nateraw. Survey Lung Cancer dataset [Data set]. Hugging Face. <https://huggingface.co/datasets/nateraw/lung-cancer>

- [9] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273-297. <https://doi.org/10.1007/BF00994018>
- [10] Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [11] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- [12] Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27. <https://doi.org/10.1109/TIT.1967.1053964>
- [13] Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- [14] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- [15] Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- [16] Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1-3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2)
- [17] Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC.
- [18] McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12, 153-157. <https://doi.org/10.1007/BF02295996>
- [19] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- [20] Eassa, M., & Abdelhafeez, A. (2026). Q-LungBench: Reproducibility code and results. GitHub repository. <https://github.com/ahmadaelhafeez-lang/Q-LungBench>

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.