Article

Combining Low-Dose Computer-Tomography-Based Radiomics and Serum Metabolomics for Diagnosis of Malignant Nodules in Participants of Lung Cancer Screening Studies

Joanna Zyla [1,†], Michal Marczyk [1,2,*], Wojciech Prazuch [1,†], Magdalena Sitkiewicz [3], Agata Durawa [3], Malgorzata Jelitto [4], Katarzyna Dziadziuszko [4], Karol Jelonek [5], Agata Kurczyk [6], Edyta Szurowska [4], Witold Rzyman [3], Piotr Widłak [4] and Joanna Polanska [1]

[1] Department of Data Science and Engineering, Silesian University of Technology, 44-100 Gliwice, Poland; joanna.zyla@polsl.pl (J.Z.); wojciechprazuch3@gmail.com (W.P.); joanna.polanska@polsl.pl (J.P.)
[2] Yale Cancer Center, Yale School of Medicine, New Haven, CT 06510, USA
[3] Department of Thoracic Surgery, Medical University of Gdansk, 80-210 Gdansk, Poland; magdalena.sitkiewicz@gumed.edu.pl (M.S.); agata.durawa@gumed.edu.pl (A.D.); wrzyman@gumed.edu.pl (W.R.)
[4] 2nd Department of Radiology, Medical University of Gdansk, 80-210 Gdansk, Poland; jelitto@gumed.edu.pl (M.J.); katarzyna.dziadziuszko@gumed.edu.pl (K.D.); eszurowska@gumed.edu.pl (E.S.); piotr.widlak@gumed.edu.pl (P.W.)
[5] Center for Translational Research and Molecular Biology of Cancer, Maria Sklodowska-Curie National Research Institute of Oncology, Gliwice Branch, 44-100 Gliwice, Poland; karol.jelonek@gliwice.nio.gov.pl
[6] Department of Biostatistics and Bioinformatics, Maria Sklodowska-Curie National Research Institute of Oncology, Gliwice Branch, 44-100 Gliwice, Poland; agata.kurczyk@gliwice.nio.gov.pl
*   Correspondence: michal.marczyk@polsl.pl
†   These authors contributed equally to this work.

Abstract: Radiomics is an emerging approach to support the diagnosis of pulmonary nodules detected via low-dose computed tomography lung cancer screening. Serum metabolome is a promising source of auxiliary biomarkers that could help enhance the precision of lung cancer diagnosis in CT-based screening. Thus, we aimed to verify whether the combination of these two techniques, which provides local/morphological and systemic/molecular features of disease at the same time, increases the performance of lung cancer classification models. The collected cohort consists of 1086 patients with radiomic and 246 patients with serum metabolomic evaluations. Different machine learning techniques, i.e., random forest and logistic regression were applied for each omics. Next, model predictions were combined with various integration methods to create a final model. The best single omics models were characterized by an AUC of 83% in radiomics and 60% in serum metabolomics. The model integration only slightly increased the performance of the combined model (AUC equal to 85%), which was not statistically significant. We concluded that radiomics itself has a good ability to discriminate lung cancer from benign lesions. However, additional research is needed to test whether its combination with other molecular assessments would further improve the diagnosis of screening-detected lung nodules.

Keywords: classification models; integration; early detection; lung cancer; screening study

1. Introduction

Lung cancer holds the unenviable position of being the leading cause of cancer-related fatalities in both men and women. It accounts for 21% of all cancer deaths among men, followed by prostate (11%) and colon and rectum cancers (9%). In women, it represents 21% of all cancer deaths, followed by breast (15%) and colon and rectum cancers (8%) [1]. In the year 2020, the global landscape witnessed a staggering 2.2 million new cases of lung cancer, resulting in 1.8 million fatalities attributed to this serious disease [2]. Notably,

in Poland, 63 individuals succumb to lung cancer each day, ranking high incidence in Europe. Approximately, 81% of lung cancer deaths result from direct cigarette smoking, with indirect smoking and air pollution being the following contributing factors. In Poland, the population of smokers is estimated at a substantial 8 million individuals. The low survival rate for lung cancer is mainly related to late diagnosis. An asymptomatic course of disease progression causes a low rate of early-stage detection. In the case of symptoms, long-term smoking patients usually develop chronic symptoms similar to early signs of lung cancer earlier but often disregard them if they are not severe or present at all. Moreover, patients' beliefs and worries about health changes that may indicate lung cancer appear to play a crucial part in the delay in diagnosis. Late diagnosis and oncological therapy often fail to cure the patient. Still, they could do so in the earlier stages of the disease, for which the survival rate is much higher [3].

The diagnosis of early lung cancer in the real world typically initiates with X-ray chest radiography or computed tomography (CT) performed for other reasons. Nowadays, the systemic approach to the detection of early-stage lung cancer has been implemented in a few countries based on the results of two pivotal randomized clinical trials: NLST and NELSON. These clinical trials provided incontrovertible evidence of reducing lung-cancer-related mortality by 20% and 26%, respectively, with the implementation of screening programs involving low-dose computed tomography (LDCT), specifically targeting high-risk groups, such as tobacco smokers [4,5]. LDCT-based lung cancer screening leads to thousands of images with a substantial number of lung nodules that should be appropriately categorized. Hence, radiologists play a crucial role in identifying nodules or tumors within the lung parenchyma and predicting the risk of malignancy. If any nodule displays malignant features, the process of invasive lung cancer diagnosis begins.

Lung cancer screening has been established to enhance the rate of early lung cancer detection in cases where the disease remains asymptomatic during its progression. These attempts target a high-risk demographic comprising middle-aged and elderly individuals with a history of long-term smoking. Ideally, lung cancer screening rounds should occur annually, which places a significant burden on current radiological resources. The evaluation of image-based screening for detecting subtle abnormalities can be a complex and time-consuming process. Moreover, a significant proportion of false positive results of CT-based tests affect lung cancer diagnosis. It is assumed that the diagnostic accuracy of cancer detection could be increased by supplementing low-dose CT imaging with additional diagnostic tests, particularly molecular biomarkers [6,7]. Among the hypothetical biomarkers of early lung cancer that could complement CT-based diagnosis are different molecular and cellular components of blood [8–11]. Metabolites present in blood are promising candidates for biomarkers since they could be potentially detected through "liquid biopsy" [12]. For example, choline-containing phospholipids and sphingolipids are serum/plasma components that discriminate between lung cancer patients and healthy individuals [13–15]. Recently, we performed a metabolomics study to search for serum metabolites that differentiated three groups of lung cancer screening participants: patients with screen-detected lung cancer, individuals with benign pulmonary nodules, and those without any lung alterations. However, despite several specific compounds having significant differences among compared groups, the low accuracy of classification models was observed (AUC = 60%) due to substantial heterogeneity in the levels of analyzed metabolites [16].

Radiomics allows the extraction of a comprehensive set of features from an LDCT image for automated cancer detection and the diagnosis of malignant lesions [17]. We hypothesized that the combination of disease-related features observed at the systemic level (i.e., the features of serum metabolome) and a depiction of local pathological changes (i.e., the features of LDCT images) would enhance the precision of lung cancer classification models. To prove our concept, we developed a method for differentiating between benign and malignant nodules detected in lung cancer screening participants by combining the results of LDCT and metabolomic modalities. We gathered data from two different screening

cohorts and tested two machine learning models and several integration algorithms to find the best solution.

2. Materials and Methods

2.1. Study Subject

Material included in this study was collected during two lung cancer screening programs performed by the Medical University of Gdansk in the years 2009–2011 (PPPBWWRP—Pomorski Pilotażowy Program Badań Wczesnego Wykrywania Raka Płuca) and 2016–2018 (MOLTEST-BIS) [3,16]. These programs enrolled more than 14 thousand participants and offered LDCT examinations for current or former smokers with at least a 20-pack-year history, aged from 50 to 75 years. This report involves two groups of participants of the project: (i) individuals with CT-detected lung nodules that were confirmed benign via histopathology, further marked as benign, and (ii) patients who were ultimately diagnosed with lung cancer, further marked as malignant. For those patients, two types of measurements (modalities) were collected: (i) radiomic characteristics from LDCT scans (1086 participants of either PPPBWWRP or MOLTEST-BIS dataset), and (ii) serum metabolome profiles (246 participants of the MOLTEST-BIS cohort). Several regions of suspicious elements from the LDCT scan were collected for one of the patients. The basic characteristics of the analyzed groups are presented in Table 1. Studies were approved by the appropriate Ethics Committees (the Medical University of Gdansk, approval nos. NKEBN/42/2009 and NKBBN/376/2014), and all participants provided informed consent indicating their voluntary participation in the project and provision of blood samples for future research.

Table 1. Basic characteristics of the analyzed population. NA means that the information was not available. Since a patient could have benign and malignant nodules at the same time, for radiomics data, n represents the number of patients with at least one nodule of a given type.

	Radiomics	Metabolomics
Benign nodules		
n	994 (75%)	123 (50%)
Screening program		
PPPBWWRP	906 (91%)	0 (0%)
MOLTEST BIS	88 (9%)	123 (100%)
Sex: male/female	445/549 (45%/55%)	66/57 (54%/46%)
Age years: (median)	NA	51–79 (67)
Smoking pack-year: range (median)	NA	26–133 (43)
Malignant nodules (lung cancer)		
n	331 (25%)	123 (50%)
Screening program		
PPPBWWRP	258 (78%)	0 (0%)
MOLTEST BIS	73 (22%)	123 (100%)
Sex: male/female	136/195 (41%/59%)	67/56 (54%/46%)
Age years: range (median)	NA	53–79 (67)
Smoking pack-year: range (median)	NA	24–138 (48)

Table 1. *Cont.*

	Radiomics	Metabolomics
Clinical stage:		
IA	NA	49
IB	NA	10
IIA	NA	9
IIB	NA	10
IIIA	NA	17
IIIB	NA	7
IVA	NA	16
IVB	NA	5

2.2. Metabolomic Data

The detailed procedure of data collection, preparation, and preprocessing is described elsewhere [16]. Briefly, for serum samples isolated from the peripheral blood of MOLTEST-BIS participants, the measures were obtained via a high-resolution mass spectrometry assay using an Absolute IDQ p400 HR kit (test plates in the 96-well format; Biocrates Life Sciences AG, Innsbruck, Austria) according to the manufacturer's protocol. For obtained measurements, batch correction and missing data imputation were performed as described elsewhere [16]. Finally, concentrations of 259 specific metabolites (or lipid isomer groups) and aggregated concentrations of different metabolite classes: acylcarnitines, amino acids, biogenic amines, glycerophospholipids, sphingolipids, cholesterol esters (CEs), glycerides, triglycerides (TGs), diglycerides (DGs), phosphatidylcholines (PCs), lysophosphatidyl-cholines (LPCs), and total lipids were analyzed; this resulted in 271 metabolomics features (Supplementary Table S1).

2.3. Radiomic Data

Radiomic data for this project come from measurements performed on 1086 patients (925 from PPPBWWRP and 161 from MOLTEST-BIS). For each patient, the LDCT was performed, and for abnormalities found in the lung parenchyma, an annotation was prepared by an expert radiologist. Regions with abnormalities were categorized into the following groups: cancer, suspicious nodules, inflammation, benign nodules, lymph nodes, fibrosis, and calcification (one patient could have different pathologies classified in different groups). Cancer, suspicious nodules, and inflammation were considered "malignant", while the remaining categories represent the "benign" group. Next, the mask of objects and their segmentations were extracted using a multi-step pre-processing and segmentation algorithm [18]. For extracted regions of the lung, the radiomics features were calculated with the usage of the PyRadiomics package version 3.0 [19]. In total, 107 radiomics features were analyzed for 5180 fragments of annotated images from analyzed patients (Supplementary Table S1). Finally, radiomics features were internally standardized (scaled and shifted) using non-parametric statistics, like median and interquartile range, calculated on benign samples within a cohort.

2.4. Univariate Analysis

Each analyzed feature from both modalities was tested due to the normality of distribution using the Shapiro–Wilk test. As the data were highly skewed, to estimate the significance of differences in analysis groups, the Mann–Whitney test was used. Moreover, each feature's biserial correlation (r_g) was calculated and treated as an effect size measure. Finally, the Benjamini–Hochberg procedure for the FDR correction was applied when necessary [20]. All statistical hypotheses were tested at the 5% significance level.

2.5. Machine Learning Sets

The classification models were constructed using two different machine learning (ML) approaches: (i) logistic regression (LR) and (ii) random forest (RF). Both ML methods were performed on the same training and test sets. The test set was extracted from the MOLTEST-BIS cohort by taking 20 benign and 20 malignant cases for which both radiomic and metabolomic data were available. Moreover, for radiomic data, the test set was expanded for additional cases from the PPPWWRP cohort. The remaining samples were gathered into a training set. A summary of the training and test set for both omics is presented in Table 2.

Table 2. Number of samples in training and test sets for ML model building.

	Radiomics	Metabolomics	Common
Train			
Benign	4569	103	68
Malignant	440	103	53
N	5009	206	121
Test			
Benign	122	20	20
Malignant	49	20	20
N	171	40	40

2.6. Logistic Regression Models

For the initially selected train set multiple random cross-validation (MRCV) was performed for logistic regression (LR), as follows: (i) data were split into training and validation (70%/30%) subsets; (ii) forward feature selection was performed on training subset with $\Delta BIC \leq 2$ as a stop criterion; (iii) an evaluation of classification parameters was conducted on the training and validation subsets; and (iv) an estimation of classification threshold was made by maximizing balanced accuracy (BAcc). The MRCV procedure was repeated 100 times. Next, feature ranking was generated as follows: (i) the features included in each model were sorted by their order of addition in the forward procedure; (ii) the proportional order was multiplied by the BAcc of the validation set at a particular fold; and (iii) the elbow technique was used to extract the most relevant features for the final LR model. Finally, the model was built on the entire training set using selected features and evaluated on the test set.

2.7. Random Forest Models

Random forest classifier was implemented using caret R package version 6.0-94 [21] with sample weighting to decrease the effect of class imbalance. Two RF model parameters were tested: (i) Mtry—number of features sampled at each tree split (from 5 to 30); (ii) Ntree—number of trees in a forest (100, 500, 1000, 2000). MRCV procedure was applied, as follows: (i) data were split into training and validation (80%/20%) subsets; (ii) RF model was fit on training data; (iii) estimation of variable importance was made (for each tree, the prediction accuracy on the out-of-bag portion of the data is recorded; then, the same is conducted after permuting each predictor variable; the difference between the two accuracies is then averaged over all trees and normalized via the standard error); and (iv) an estimation of classification threshold is made by maximizing balanced accuracy. The MRCV procedure was repeated 100 times. Next, feature ranking was generated, as follows: (i) for each feature, the average variable importance score was calculated; (ii) the elbow technique was used to extract the most relevant features for the final RF model. Finally, the model was built on the entire training set using selected features and evaluated on the test set.

2.8. Machine Learning Result Integration

Several approaches were applied to integrate results from both platforms. The first one was based on statistical integration proposed by Stouffer [22]. It was used for both classification probabilities of test sets as well as classification thresholds. Additionally, several common methods were tested including: (i) mean value; (ii) maximum value; and (iii) product of two probabilities. All methods are described in detail elsewhere [23]. Similarly, both classification probabilities and classification thresholds were integrated using the same method.

3. Results

To construct classification models, we used two different cohorts, PPPBWWRP and MOLTEST-BIS, and two types of measured features, radiomic features from LDCT images and concentrations of serum metabolites (metabolomic features). We cleaned the datasets by removing missing values and the normalization of the data (see Section 2). For radiomic features, significant differences were not observed between cohorts after data normalization (Figure 1). Better separation between benign and malignant cases was noted using radiomic features compared to metabolomic features, but in general, these two classes are not separated. The PPPBWWRP radiomic dataset revealed three clusters of data points, of which the largest one included both benign and malignant cases (Figure 1). A refined investigation revealed that one of the smaller clusters included mostly calcified nodules that resulted from historical infections or physical damages (Supplementary Figure S1), while the smallest one consists of all types of benign nodules and might represent unknown technical artifacts.
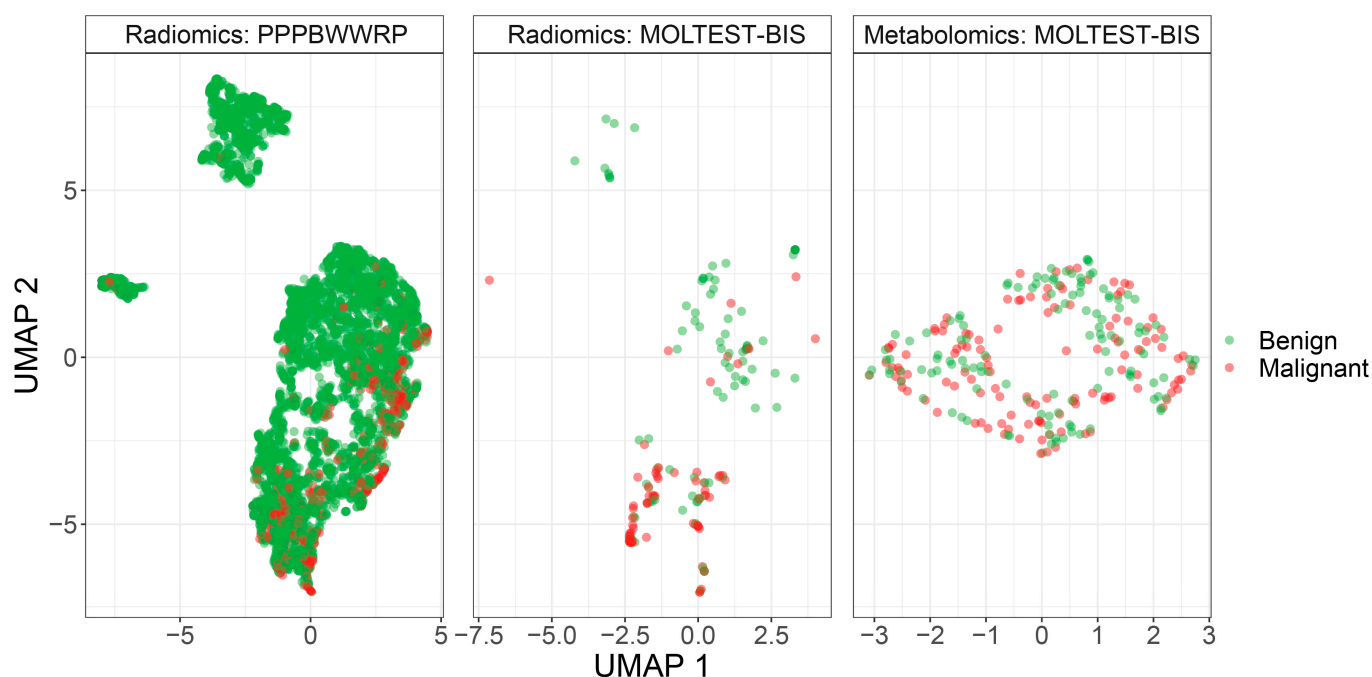

Figure 1. UMAP visualization of patients' clustering using either radiomic or metabolomic modalities.

3.1. Univariate Analysis of Metabolomic and Radiomic Studies

Many features measured in both modalities were highly skewed, showing non-normal distribution (Supplementary Table S1, Supplementary Figure S2), so we used non-parametric methods for univariate analysis to compare benign and malignant cases within each modality. We ran the analysis using all available samples in each modality. After applying multiple testing corrections, we found 94 statistically significant (FDR < 0.05) features in the radiomic data, from which 44 were downregulated and 50 upregulated. However, no statistically significant features were identified in the metabolomic dataset.

The mean effect size measured for top differentially regulated features was also higher in radiomic data than in metabolomic data (0.6 vs. 0.26; examples of differentially regulated features are presented in Figure 2). Many features showed similar patterns of level difference, which is due to the high correlation between them within modalities (Supplementary Figure S3).
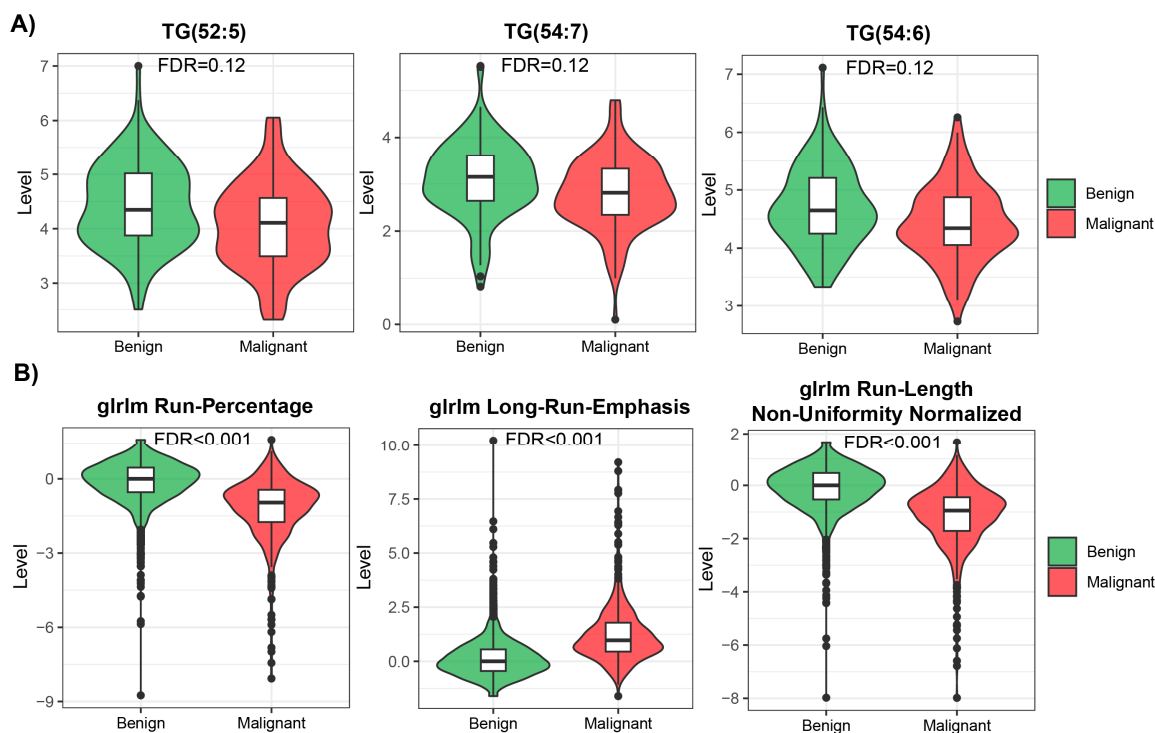

Figure 2. Top 3 most differentially changed features in metabolomic (**A**) and radiomic (**B**) modalities.

3.2. Development of Machine Learning Models

Before modeling, we removed highly correlated features in each modality. The MRCV procedure used for parameter tuning and feature selection (Supplementary Figures S4 and S5) resulted in the following models: (i) metabolomics—11 features for LR and 14 for RF; (ii) radiomics—11 features for LR and 8 for RF. Only three features were common between both ML models in metabolomics mode (PC(41:5), PC-O(42:6), and PC(42:7)) and three in radiomics mode (glcm InverseVariance, shape Flatness, and glcm Id). As can be observed in Supplementary Figures S6 and S7, as well as Supplementary Table S2, the LR approach has good classification performance in training and test sets for radiomic data. However, in the case of metabolomics, RF has better performance in the test set. Next, the results from metabolomic and radiomic modalities were integrated within each ML approach using four different methods (see Section 2). Statistical and product integrations show the best results. In both, more than half of the patients in the test set were properly classified by all models after integration (Figure 3). In most cases, integration decreased the number of false positive and negative findings, giving superior results compared to a model based on only one modality.

Looking at the model performance indices on the test set after applying the estimated classification thresholds, we observed diverse results. The RF model was better than the LR model for metabolomic data; however, it was worse for radiomics mode and after statistical and product integration (Table 3). The overall performance was moderate with F1 scores ranging from 0.55 to 0.8 and AUCs ranging from 55.5% to 84.8%. ROC curve analysis showed that there is a potential to tune threshold values to meet other specific goals of a model (Figure 4). The ROC curves for the other tested integration methods are presented in Supplementary Figure S8. Finally, the results of the training set showed that models were not significantly overtrained (Supplementary Table S2, Supplementary Figure S9).

Integrating classification results from two modalities slightly increased model performance in comparison to the basic radiomics model (e.g., F1 and BAcc), regardless of the integration method used (Table 3; Figure 4). However, this increase was not statistically significant.
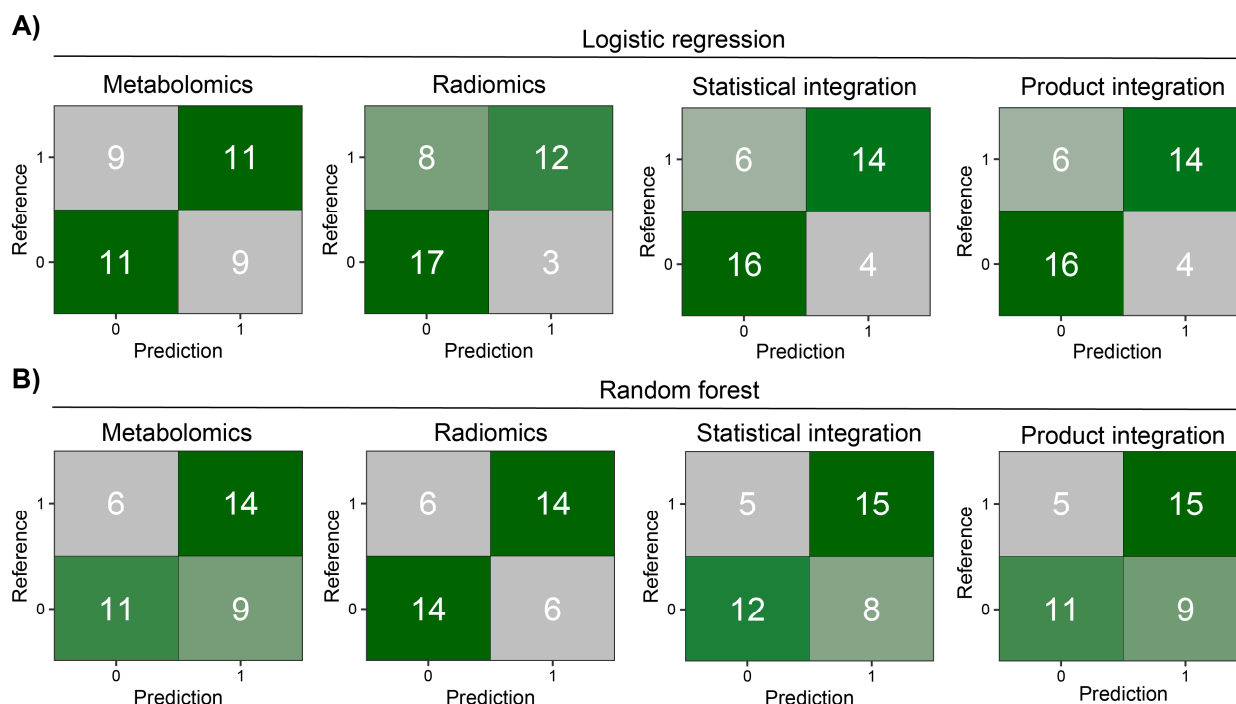

Figure 3. Heatmaps of confusion matrices for prediction models on the MOLTEST-BIS test set for each modality and after integration. (**A**) Results of logistic regression models. (**B**) Results of random forest models.

Table 3. Results for ML models on MOLTEST-BIS test set. The bold value shows the superiority of a particular metric and the set of data between ML approaches. LR—logistic regression, RF—random forest, CI—95% confidence interval.

| Metric	Radiomics		Metabolomics		Statistical Integration		Product Integration	
	LR	RF	LR	RF	LR	RF	LR	RF
Sensitivity	0.60	**0.70**	0.55	**0.70**	0.70	**0.75**	**0.78**	0.63
Specificity	**0.85**	0.70	**0.55**	**0.55**	**0.80**	0.60	0.73	0.69
PPV	**0.80**	0.70	0.55	**0.61**	**0.78**	0.65	0.70	**0.75**
NPV	0.68	**0.70**	0.55	**0.65**	**0.73**	0.71	**0.80**	0.55
F1	0.69	**0.70**	0.55	**0.65**	**0.74**	0.70	**0.74**	0.68
Balanced accuracy	**0.73**	0.70	0.55	**0.63**	**0.75**	0.68	**0.75**	0.66
AUC (%)	**83.0**	70.3	55.5	**60.3**	**83.0**	73.0	**84.8**	71.5
AUC 95% CI	**70–96**	53–87	38–74	**42–79**	**70–96**	56–89	**73–97**	55–88
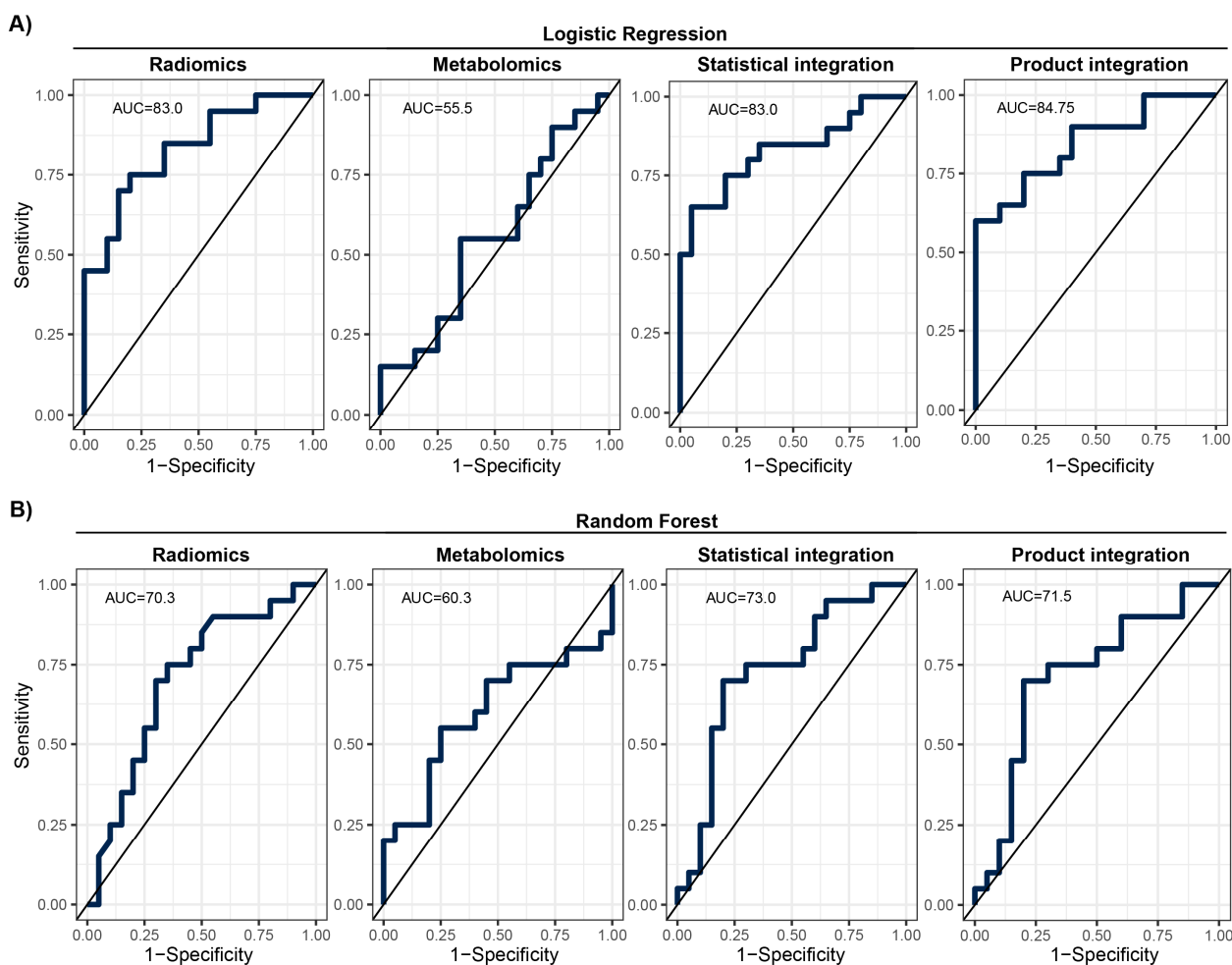
Figure 4. ROC with given AUC for prediction models on the MOLTEST-BIS test set for each modality and after integration. (**A**) Results of logistic regression models. (**B**) Results of random forest models.

4. Discussion

We used two different modalities to build the prediction model for early lung screening support: (i) radiomic characteristics from LDCT scans collected in a large cohort, and (ii) serum metabolome profiles collected in a study with a smaller sample size. In the univariate analysis, a pool of significant features was found for radiomics. In contrast, for metabolomics, none of the investigated serum metabolome features were significant (FDR < 0.05). This indicates that LDCT has a better ability to distinguish lung cancer in the presented study on a single feature level. Yet, the size of the cohort used for radiomics was much bigger than for metabolomics, which can impact the observed result. Out of the significant radiomic features, run percentage (RP) can be primarily distinguished, which measures the coarseness of the texture by taking the ratio of the number of runs and number of voxels in the ROI. A higher value indicates a finer texture, and in Figure 2B (first graph), we can observe much lower values for malignant cases. The second highlighted feature is long run emphasis (LRE), which measures the distribution of long run lengths, and greater value indicates longer run lengths and more coarse structural textures. For malignant cases, we observed a higher value of LRE, as expected (Figure 2B, second graph). Next, we constructed two separate models for both modalities. Again, the model built using radiomic features showed better performance compared to the metabolomics-based model. The AUC for the best radiomics model in the test set was 83% (LR model), while for metabolomics, it was only 60.3% (RF model). The better-performing metabolomics model (RF) included 14 features (namely, Histidine, Spermidine, PC(41:5), PC(42:7), PC(42:2),

PC(33:4), PC-O(42:6), LPC-O(16:1), AC(0:0), AC(8:1), CE(17:0), TG(51:1), TG(51:4), and TG(44:4)). Noteworthily, a reduced level of histidine was previously noted in the sera of patients with non-small-cell lung cancer [24]. Additionally, increased concentrations of spermidine were noted in serum/plasma and urine of patients with different types of malignancies including lung cancer [25]. However, for other (lipid) components, data on specific associations with lung cancer was not determined. When radiomic features were considered, better performance was observed for the LR model with the following features: glcm InverseVariance, shape Flatness, glszm ZonePercentage, ngtdm Strength, firstorder InterquartileRange, glrlm RunLengthNonUniformityNormalized, glcm MCC, glrlm LongRunLowGrayLevelEmphasis, glcm Id, glszm LargeAreaLowGrayLevelEmphasis, and gldm LargeDependenceHighGrayLevelEmphasis. Several features were previously reported in other studies, e.g., LargeAreaLowGrayLevelEmphasis shows effective discrimination of lung cancer from tuberculosis with AUC 92% [26]. Other examples are ngtdm Strength and glszm ZonePercentage, which showed the effectiveness of the response to immunotherapy for non-small-cell lung cancer [27]. Finally, we integrated predictions from both modalities. The product integration on LR shows slightly better performance compared to a single radiomics model, with an AUC of 84.75%. Moreover, the observed NPV was the highest (80%). Yet, the highest PPV was observed for radiomic features and the LR model.

To summarize, our results indicate that combining the outcomes of the machine learning models based on two different modalities—LDCT radiomics and serum metabolomics—slightly increases the potential of available methods to build diagnostic tools that could discriminate benign and malignant nodules detected in participants of lung cancer screening programs. However, the preliminary results of our pilot study must be extended and validated using larger cohorts from other screening studies. Hopefully, authors of future screening programs will perform metabolic and radiomic analyses and share their data. Moreover, a combination of radiomics with other molecular or genomic signatures may result in more promising outcomes. For example, the combination of miRNA classifiers with radiomic features resulted in the increased performance of pancreatic cancer diagnosis [28]. Also, a combination of radiomics with clinical characteristics may further enhance the performance of classification models, which was previously reported for the discrimination between pneumonia-like lung cancer from pulmonary inflammatory lesions [29].

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biom14010044/s1, Table S1. Univariate analysis results. Table S2. Results for ML models on train and test set (without removal of MOLTEST-BIS from test set). The bold value shows the superiority of a particular metric and set of data between models; LR—logistic regression, RF—random forest. Figure S1. UMAP visualization of all patients in radiomics with detailed annotated regions. Figure S2. Distribution of signal skewness for each omics and investigated group. Panel A shows metabolomics while panel B shows radiomics results. High values on the *x*-axis represent right skewness, while low values represent left skewness. Figure S3. Spearman correlation between features in metabolomics (A) and radiomics (B) modalities. Figure S4. Elbow plot of feature ranking for logistic regression approach. Panels A and B show results for metabolomics and radiomics respectively. Pale blue marks features that are used in the final classifier. Figure S5. Elbow plot of feature ranking for random forest approach. Panels A and B show results for metabolomics and radiomics respectively. Pale blue marks features that are used in the final classifier. Figure S6. Heatmaps of confusion matrices for LR models on the train and full test sets for metabolomics (A) and radiomics (B). Figure S7. Heatmaps of confusion matrices for RF models on the train and full test sets for metabolomics (A) and radiomics (B). Figure S8. ROC with given AUC for test set result for other integration approaches. Panel A shows results for logistic regression; Panel B shows results for random forest. Figure S9. ROC with given AUC for the training set, test set, and each modality. Panel A shows results for logistic regression; Panel B shows results for random forest.

Author Contributions: Conceptualization, J.P., W.R. and P.W.; methodology, J.Z., W.P., M.M. and J.P.; software, J.Z., W.P., M.M. and J.P.; validation, J.Z., M.M. and J.P.; formal analysis, J.Z., W.P. and M.M.; investigation, J.Z., W.P., K.J., A.K. and M.M.; resources, A.D., M.J., K.D. and K.J.; data

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the appropriate Ethics Committee: Medical University of Gdańsk-approvals no. NKEBN/42/2009 and NKBBN/376/2014.

Informed Consent Statement: Written informed consent has been obtained from all participating patients and controls involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to privacy and ethical reasons.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Schabath, M.B.; Cote, M.L. Cancer Progress and Priorities: Lung Cancer. *Cancer Epidemiol. Biomark. Prev.* **2019**, *28*, 1563–1579. [CrossRef] [PubMed]
2. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
3. Ostrowski, M.; Marjański, T.; Dziedzic, R.; Jelitto-Górska, M.; Dziadziuszko, K.; Szurowska, E.; Dziadziuszko, R.; Rzyman, W. Ten years of experience in lung cancer screening in Gdańsk, Poland: A comparative study of the evaluation and surgical treatment of 14,200 participants of 2 lung cancer screening programmes. *Interact. CardioVascular Thorac. Surg.* **2019**, *29*, 266–274. [CrossRef] [PubMed]
4. The National Lung Skrining Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Skrining. *N. Engl. J. Med.* **2011**, *365*, 395–409. [CrossRef]
5. de Koning, H.J.; van der Aalst, C.M.; de Jong, P.A.; Scholten, E.T.; Nackaerts, K.; Heuvelmans, M.A.; Lammers, J.W.J.; Weenink, C.; Yousaf-Khan, U.; Horeweg, N.; et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N. Engl. J. Med.* **2020**, *82*, 503–513. [CrossRef]
6. Priola, A.M.; Priola, S.M.; Giaj-Levra, M.; Basso, E.; Veltri, A.; Fava, C.; Cardinale, L. Clinical implications and added costs of incidental findings in an early detection study of lung cancer by using low-dose spiral computed tomography. *Clin. Lung Cancer* **2013**, *14*, 139–148. [CrossRef]
7. Atwater, T.; Massion, P.P. Biomarkers of risk to develop lung cancer in the new screening era. *Ann. Transl. Med.* **2016**, *4*, 158. [CrossRef]
8. Hassanein, M.; Callison, J.C.; Callaway-Lane, C.; Aldrich, M.C.; Grogan, E.L.; Massion, P.P. The state of molecular biomarkers for the early detection of lung cancer. *Cancer Prev. Res.* **2012**, *5*, 992–1006. [CrossRef]
9. Sozzi, G.; Boeri, M. Potential biomarkers for lung cancer screening. *Transl. Lung Cancer Res.* **2014**, *3*, 139–148.
10. Chu, G.C.W.; Lazare, K.; Sullivan, F. Serum and blood based biomarkers for lung cancer screening: A systematic review. *BMC Cancer* **2018**, *18*, 181. [CrossRef]
11. Ostrin, E.J.; Sidransky, D.; Spira, A.; Hanash, S.M. Biomarkers for lung cancer screening and detection. *Cancer Epidemiol. Biomark. Prev.* **2020**, *29*, 2411–2415. [CrossRef] [PubMed]
12. Spratlin, J.L.; Serkova, N.J.; Eckhardt, S.G. Clinical applications of metabolomics in oncology: A review. *Clin. Cancer Res.* **2009**, *15*, 431–440. [CrossRef] [PubMed]
13. Rocha, C.M.; Carrola, J.; Barros, A.S.; Gil, A.M.; Goodfellow, B.; Carreira, I.M.; Bernardo, J.; Gomes, A.; de Sousa, V.M.L.; Carvalho, L.; et al. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of blood plasma. *J. Proteome Res.* **2011**, *10*, 4314–4324. [CrossRef] [PubMed]
14. Guo, Y.; Wang, X.; Qiu, L.; Qin, X.; Liu, H.; Wang, Y.; Li, F.; Wang, X.; Chen, G.; Song, G.; et al. Probing gender-specific lipid metabolites and diagnostic biomarkers for lung cancer using Fourier transform ion cyclotron resonance mass spectrometry. *Clin. Chim. Acta* **2012**, *414*, 135–141. [CrossRef] [PubMed]

15. Chen, Y.; Ma, Z.; Min, L.; Li, H.; Wang, B.; Zhong, J.; Dai, L. Biomarker identification and pathway analysis by serum metabolomics of lung cancer. *BioMed Res. Int.* **2015**, *2015*, 183624. [CrossRef] [PubMed]

16. Widłak, P.; Jelonek, K.; Kurczyk, A.; Żyła, J.; Sitkiewicz, M.; Bottoni, E.; Veronesi, G.; Polańska, J.; Rzyman, W. Serum Metabolite Profiles in Participants of Lung Cancer Screening Study; Comparison of Two Independent Cohorts. *Cancers* **2021**, *13*, 2714. [CrossRef] [PubMed]

17. Binczyk, F.; Prazuch, W.; Bozek, P.; Polanska, J. Radiomics and artificial intelligence in lung cancer screening. *Transl. Lung Cancer Res.* **2021**, *10*, 1186–1199. [CrossRef]

18. Prazuch, W.; Jelitto-Gorska, M.; Durawa, A.; Dziadziuszko, K.; Polanska, J. Radiomic-Based Lung Nodule Classification in Low-Dose Computed Tomography. In *International Work-Conference on Bioinformatics and Biomedical Engineering*; Springer: Cham, Switzerland, 2022.

19. Van Griethuysen, J.J.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.; Fillion-Robin, J.C.; Pieper, S.; Aerts, H.J. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [CrossRef]

20. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [CrossRef]

21. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [CrossRef]

22. Stouffer, S.A.; Suchman, E.A.; DeVinney, L.C.; Star, S.A.; Williams, R.M., Jr. *The American Soldier: Adjustment during Army Life. (Studies in Social Psychology in World War II)*; Princeton University Press: Princeton, NJ, USA, 1949; Volume 1, pp. 87–90.

23. Mohandes, M.; Deriche, M.; Saliyu, O.A. Classifiers combination techniques: A comprehensive review. *IEEE Access* **2018**, *6*, 19626–19639. [CrossRef]

24. Klupczynska, A.; Dereziński, P.; Garrett, T.J.; Rubio, V.Y.; Dyszkiewicz, W.; Kasprzyk, M.; Kokot, Z.J. Study of early stage non-small-cell lung cancer using Orbitrap-based global serum metabolomics. *J. Cancer Res. Clin. Oncol.* **2017**, *143*, 649–659. [CrossRef] [PubMed]

25. Tse, R.T.H.; Wong, C.Y.P.; Chiu, P.K.F.; Ng, C.F. The Potential Role of Spermine and Its Acetylated Derivative in Human Malignancies. *Int. J. Mol. Sci.* **2022**, *23*, 1258. [CrossRef] [PubMed]

26. Thattaamuriyil Padmakumari, L.; Guido, G.; Caruso, D.; Nacci, I.; Del Gaudio, A.; Zerunian, M.; Polici, M.; Gopalakrishnan, R.; Sayed Mohamed, A.K.; De Santis, D.; et al. The role of chest CT radiomics in diagnosis of lung cancer or tuberculosis: A pilot study. *Diagnostics* **2022**, *12*, 739. [CrossRef] [PubMed]

27. Barabino, E.; Rossi, G.; Pamparino, S.; Fiannacca, M.; Caprioli, S.; Fedeli, A.; Zullo, L.; Vagge, S.; Cittadini, G.; Genova, C. Exploring response to immunotherapy in non-small cell lung cancer using delta-radiomics. *Cancers* **2022**, *14*, 350. [CrossRef]

28. Permuth, J.B.; Choi, J.; Balarunathan, Y.; Kim, J.; Chen, D.T.; Chen, L.; Orcutt, S.; Doepker, M.P.; Gage, K.; Zhang, G.; et al. Combining radiomic features with a miRNA classifier may improve prediction of malignant pathology for pancreatic intraductal papillary mucinous neoplasms. *Oncotarget* **2016**, *7*, 85785. [CrossRef]

29. Gong, J.W.; Zhang, Z.; Luo, T.Y.; Huang, X.T.; Zhu, C.N.; Lv, J.W.; Li, Q. Combined model of radiomics, clinical, and imaging features for differentiating focal pneumonia-like lung cancer from pulmonary inflammatory lesions: An exploratory study. *BMC Med. Imaging* **2022**, *22*, 98. [CrossRef]