PhD Programme in Biotechnologies in

Translational Medicine

Department of Experimental Medicine (DIMES)

XXXVIII Cycle

University of Genova

The role of radiomics and delta-radiomic analysis for the prediction of prognostic and predictive factors in non-small cell lung cancer.

Head of Programme: Prof. Sveva Bollini

Thesis Supervisor: Prof. Carlo Martinoli

Thesis Assistant Supervisors: Prof. Carlo Genova, Dr. Giovanni Rossi

Candidate: Dr. Silvia Pamparino

ABSTRACT

Background:

Lung cancer is the leading cause of cancer-related death worldwide, despite the increase in therapeutic options and screening programs. Immunotherapy has significantly transformed the treatment of non-small cell lung cancer (NSCLC), however the long-term benefit is limited to a small subset of patients. However, apart from PD-L1 tissue expression—which has been repeatedly questioned—there are currently no reliable tools to predict treatment response.

Study design and methods:

This study included patients with advanced NSCLC from three different hospitals treated with anti-PD1 monotherapy as first or second line. All patients underwent a pre-treatment CT scan, and the lesion targets were semi-automatically segmented. Radiomic features were extracted and used to develop machine learning models to predict response to immunotherapy, labeled dichotomously as progressor or non-progressor.

Results:

A total of 125 patients were enrolled, and 165 parenchymal lung lesions and 86 lymph node lesions were considered. A support vector machine (SVM) model was developed for each lesion type, achieving an F1-score of 0.77 for parenchymal lesions and 0.86 for lymph node lesions in the internal dataset. In two external validation cohorts, the model achieved an F1-score of 0.69 for parenchymal lesions and 0.77 for lymph nodes, demonstrating strong generalizability across different scanners, protocols, and readers.

Conclusions:

These radiomics-based machine learning models showed good reproducibility

across external datasets and outperformed tissue-assessed PD-L1 expression in response prediction accuracy. These findings support the growing role of radiomics as a reliable and feasible tool in clinical decision making for NSCLC immunotherapy.

OBJECTIVES AND GANTT CHART

1. The primary objective of this project is to analyze radiomic features and to define the role of radiomics and delta-radiomics in predicting prognostic and predictive factors in non-small cell lung cancer (NSCLC). Specifically, the goal is to develop an artificial intelligence algorithm capable of predicting the immunotherapy response of individual lesions by analyzing features extracted from baseline CT images, prior to the initiation of treatment ("lesion-level approach").
2. To enhance the reproducibility of both segmentation and features extraction, ensuring increasingly rapid and standardized processes, with the ultimate aim of integrating radiomics into routine clinical practice as a reliable and user-friendly tool.
3. To establish, through the creation of continuously updated imaging databases, well-defined patient subpopulations with more accurate selection and categorization. This approach is intended to improve predictive performance and ensure robust reproducibility, particularly in predictive models derived from baseline CT imaging.
4. To integrate non-invasive approaches, namely radiomic features with circulating biomarkers from liquid biopsy, in order to achieve predictive and prognostic accuracies ranging from 80% to 90% in the validation cohort and in at least one external test set.

TASK	ACTIVITY	Year 1	Year 2	Year 3
1	Patient recruitment and database design.	[redacted]		
2	Segmentation of lesions on radiological images.	[redacted]	[redacted]	[redacted]
3	Identification and clinical validation of radiological biomarkers and their subsequent reproducibility.	[redacted]	[redacted]	[redacted]
4	Statistical analysis and evaluation of results.		[redacted]	[redacted]
5	Presentation and publication of results.			[redacted]

1. Introduction

As per the Global Cancer Observatory (GLOBOCAN) 2020 data, lung cancer is the second most commonly diagnosed cancer worldwide with 2.2 million new cases in 2020. Lung cancer is currently the leading cause of cancer-related mortality worldwide with about 1.8 million deaths per year (18.7% in 2022). NSCLC tumors are currently the third most common cancer in women (12%) and the second in men (15%).[1] Epidemiological data indicate that the main risk factor for the development of lung cancer is tobacco (cigarette smoking). Other risk factors include hydrocarbon and industrial pollution, radon, asbestos, uranium chromium, chronic obstructive pulmonary disease (or COPD), and lung fibrosis.[2]

The determination of the histological type (and, in many cases, molecular and biomarker profile of the tumor) is essential for the patient because it allows the most appropriate choice of treatment. The big family of lung cancer is traditionally divided into two primary histological subtypes: non-small cell lung cancer (NSCLC; accounting for ~ 85%) and small cell lung cancer or microcytoma (SCLC; accounting for ~ 15%). NSCLC is classified into three main histological subtypes distinguished by appearance on histology – adenocarcinoma (originating from the mucus gland cells, accounting for ~ 40-45%), squamous cell carcinoma (originating from the airway squamous epithelium, accounting for ~25-30%) and large cell carcinoma (a heterogeneous group of undifferentiated epithelium-derived tumours, accounting for ~15%), although rare cases of adenosquamous (a mixture of squamous and gland-like cells) and sarcomatoid (a mixture of carcinoma and sarcoma cells) tumours also occur. (*Figure 1*) Adenocarcinoma tends to occur in the peripheral bronchi and is the most common NSCLC histologic subtype.[1,2]

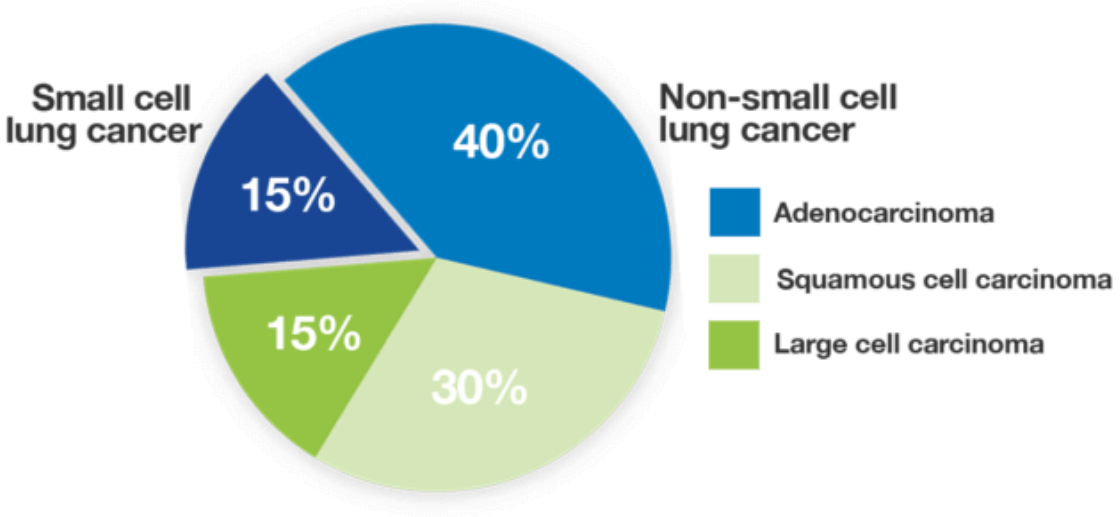


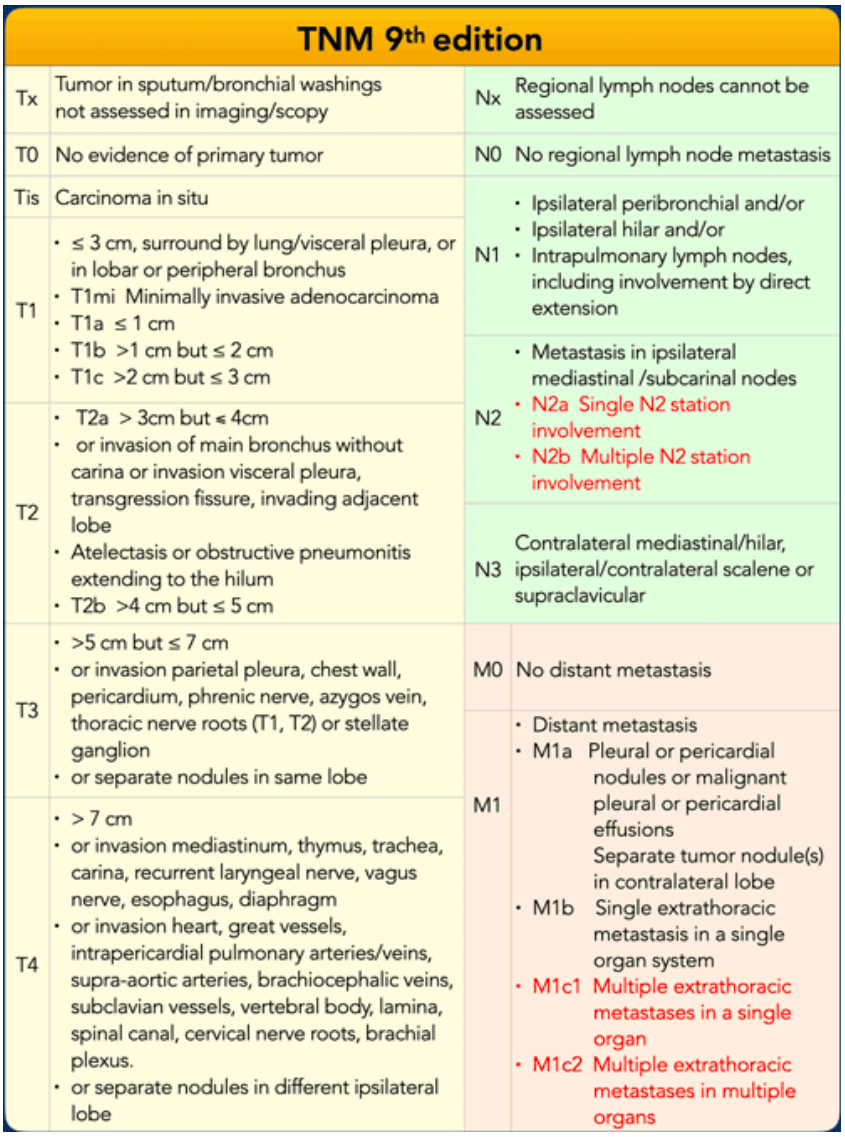
Figure 1: A broad classification of lung cancer based on histology. – Lung cancer foundation America [Types of Lung Cancer | Lung Cancer Foundation of America]

Despite the increase in therapeutic options and early detection screening programs, it remains a high-mortality cancer, being the leading cause of cancer death in both men and women, with 1.8 million deaths in 2020 alone, representing approximately 18% of all cancer mortality.[1]

Imaging is crucial in the screening, diagnosis, staging, response assessment, and surveillance of patients with lung cancer. Chest radiography remains the first-line investigation for lung cancer in primary care. The National Institute for Health and Care Excellence (NICE) lung cancer guidelines recommend chest radiography for initial evaluation in all patients, aside from those aged >40 years who have unexplained haemoptysis. Its advantages include its widespread availability, technical feasibility, simplicity, low radiation, low risk and low cost. The sensitivity and specificity of chest radiography for the detection of lung cancer are 78.3% and 97.0%, respectively, whereas for CT the sensitivity and specificity are 87.7% and 99.3%, respectively. Chest radiography only detects lung cancer in ~77–80% of cases. Therefore, CT of the thorax is used as the primary method of cancer screening, and CT is

subsequently performed if abnormalities are detected on chest radiographs. National Comprehensive Cancer Network (NCCN) and European Society of Medical Oncology (ESMO) recommend in each patient with a suspicion or a new diagnosis of lung cancer a contrast-enhanced CT scan of the chest and abdomen is required. For all patients with NSCLC who are potentially eligible for curative-intent therapy, a 2-fluoro-2-deoxy-d-glucose (FDG)-PET scan should be performed to detect occult metastases not observed with the CT scan, as FDG-PET is superior to CT scan in detecting metastases. Brain imaging is highly recommended as the incidence of brain metastases is high and this diagnosis can change the treatment plan.[3]

In NSCLC, the anatomical extent of the disease, as described by the TNM classification, is essential for the prognosis and treatment decision, and it provides a standardized description of the disease, which facilitates the clinicians in the management of the patient. The TNM classification is composed of T (representing primary tumour size), N (representing the regional pathological lymph nodes) and M (metastases). Based on these components, NSCLC can be classified into four different stages, from I (very early disease) to IV (metastatic). Developed by the International Association for the Study of Lung Cancer (IASLC) and adopted by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC), the new 9th edition of TNM classification system is effective from January 1, 2025, and it is essential that radiologists become familiar with it to incorporate the staging into their reports and discussions in multidisciplinary meetings, ensuring a common language across clinicians. In TNM-9, changes have been made mainly in the N2 and M1 subcategories based on the differences in prognostic and survival outcomes, and availability of new treatment options.[4] (*Table 1)*



TNM 9th edition

Tx	Tumor in sputum/bronchial washings not assessed in imaging/scopy	Nx	Regional lymph nodes cannot be assessed
T0	No evidence of primary tumor	N0	No regional lymph node metastasis
Tis	Carcinoma in situ	N1	• Ipsilateral peribronchial and/or • Ipsilateral hilar and/or • Intrapulmonary lymph nodes, including involvement by direct extension
T1	• ≤ 3 cm, surround by lung/visceral pleura, or in lobar or peripheral bronchus • T1mi Minimally invasive adenocarcinoma • T1a ≤ 1 cm • T1b >1 cm but ≤ 2 cm • T1c >2 cm but ≤ 3 cm	N2	• Metastasis in ipsilateral mediastinal /subcarinal nodes • N2a Single N2 station involvement • N2b Multiple N2 station involvement
T2	• T2a > 3cm but ≤ 4cm • or invasion of main bronchus without carina or invasion visceral pleura, transgression fissure, invading adjacent lobe • Atelectasis or obstructive pneumonitis extending to the hilum • T2b >4 cm but ≤ 5 cm	N3	Contralateral mediastinal/hilar, ipsilateral/contralateral scalene or supraclavicular
T3	• >5 cm but ≤ 7 cm • or invasion parietal pleura, chest wall, pericardium, phrenic nerve, azygos vein, thoracic nerve roots (T1, T2) or stellate ganglion • or separate nodules in same lobe	M0	No distant metastasis
T4	• > 7 cm • or invasion mediastinum, thymus, trachea, carina, recurrent laryngeal nerve, vagus nerve, esophagus, diaphragm • or invasion heart, great vessels, intrapericardial pulmonary arteries/veins, supra-aortic arteries, brachiocephalic veins, subclavian vessels, vertebral body, lamina, spinal canal, cervical nerve roots, brachial plexus. • or separate nodules in different ipsilateral lobe	M1	• Distant metastasis • M1a Pleural or pericardial nodules or malignant pleural or pericardial effusions Separate tumor nodule(s) in contralateral lobe • M1b Single extrathoracic metastasis in a single organ system • M1c1 Multiple extrathoracic metastases in a single organ • M1c2 Multiple extrathoracic metastases in multiple organs

Table 1: 9th edition of lung cancer TNM classification, with changes from the previous 8th edition shown in red. https://radiologyassistant.nl/chest/lung-cancer/tnm-classification-8th-edition-1

Approximately 25-30% of non-small-cell lung cancer (NSCLC) are detected in an early stage and are considered resectable, therefore amenable to surgical treatment, which permits pathologic evaluation of the resected specimen; unfortunately, approximately 70-75% of cases are unresectable at the time of the diagnosis, and the treatment choice is based on the small biopsy sample. Unresectable disease typically includes cancers that have spread to distant

organs (Stage IV) or are locally advanced with significant lymph node involvement (Stage IIIb or unresectable Stage IIIa).[4]

Treatment is based on histological and molecular characterization to offer patients increasingly "targeted" therapies, that is, specific to the characteristics of the individual tumor. *(Figure 2)* Tissue for the characterization is obtained through image-guided biopsy (CT-guided/ultrasound-guided/cone-beam CT-guided) or transbronchial biopsy (EBUS) *(Figure 3),* mainly depending on the lesion location and size. More and more rarely, in select cases, a surgical biopsy is required.[5]

Biopsy remains the gold standard for histological and molecular analysis, with some intrinsic limitations: both transthoracic and transbronchial biopsies are invasive techniques, with a non-negligible rate of complications (e.g., about 20-30% incidence of pneumothorax peri-procedural in CT-guided transthoracic biopsy) *(Figure 4),* as well as technical limitation of the procedure, such as in the case of transthoracic biopsy for central nodules, extensive emphysema, pulmonary fibrosis, and for both approaches, coagulation disorders and advanced pulmonary fibrosis stages. While central lesions in the lung parenchyma and mediastinum suspicious for malignancy are ideally targeted by bronchoscopy or EBUS transbronchial needle aspiration, peripheral pulmonary lesions can be sampled with image-guided assistance, mainly CT-guided. CT-guided transthoracic lung biopsy is performed as needle aspiration or cutting needle biopsy. Both types of CT-guided sampling show greater sensitivity in confirming lung cancer diagnosis than using transthoracic biopsy with fluoroscopy guidance.[1,5]

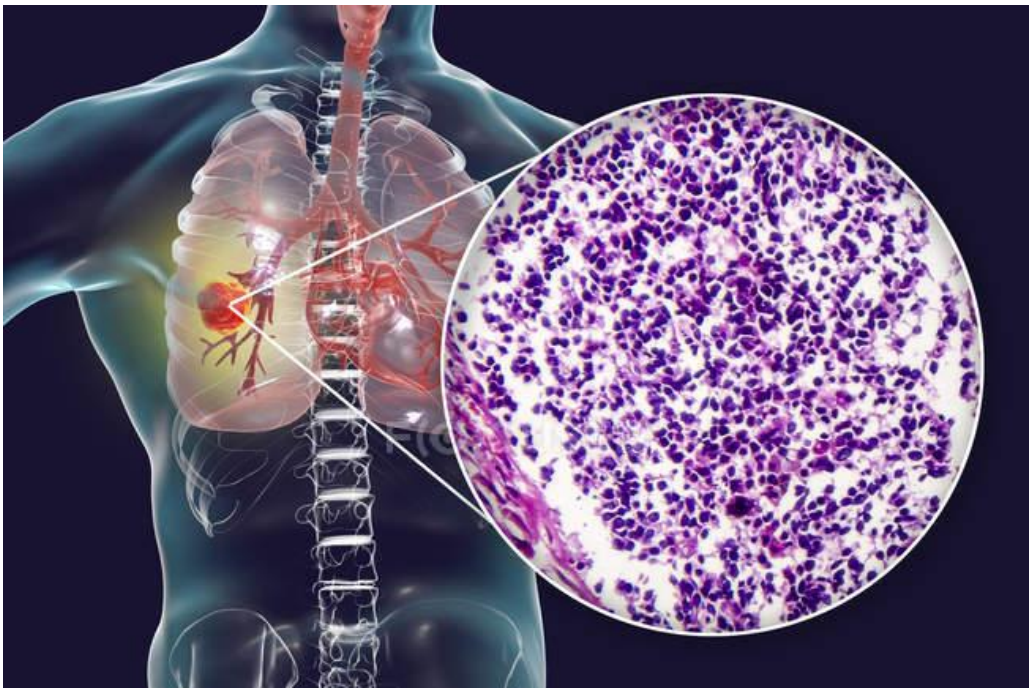


Figure 2: Histological and molecular analysis is the key to treatment decision https://www.lungcancers.eu/diagnosis/histological-diagnosis/

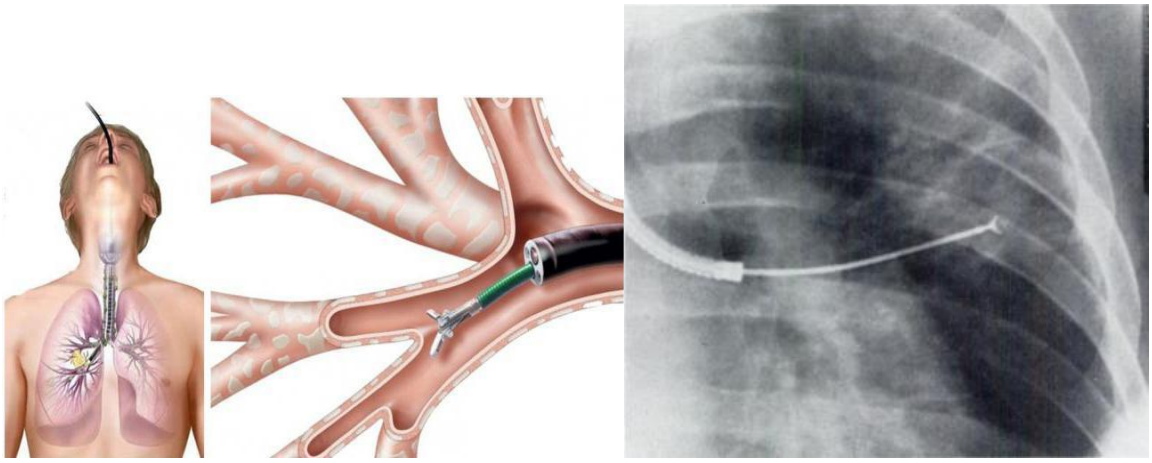


Figure 3: *Transbronchial biopsy is performed by pulmonologists and allows tissue sample mainly of center lung lesion and mediastinal lymph nodes. Modified from http://drnagarjuna.com/index.php/specialties/interventional-pulmonology/basic-bronchoscopy*

Figure 4: CT-guided lung biopsy, using a core needle biopsy, is performed by interventional radiologists and it allows tissue samples of lung mass (as shown in this picture) and nodules.

In 2015, guidelines recommended the evaluation of only two driver mutations in the specimens/biopsy samples: epidermal growth factor receptor (EGFR) mutations and anaplastic lymphoma kinase (ALK) fusions in patients NSCLC. Current guidelines recommend routine testing for the presence of multiple additional driver gene mutations or fusions (eg, ROS1, BRAF, NTRK1-3, RET, KRAS, and MET) in addition to EGFR and ALK and testing for the immune biomarker programmed death ligand 1 (PD-L1).[6]

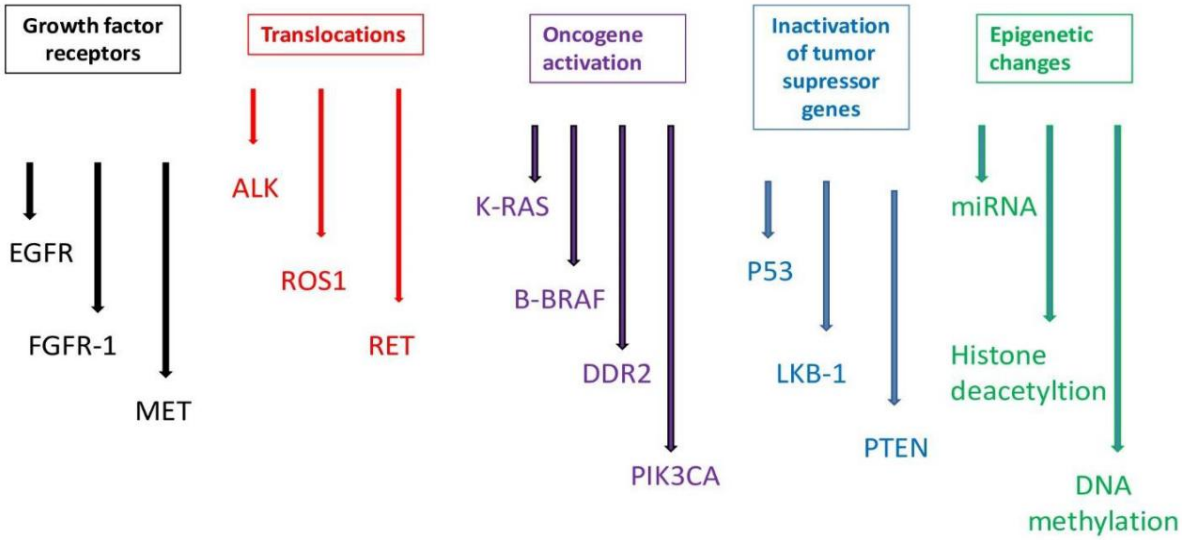


Figure 1: Most common molecular alterations involved in the pathogenesis of lung cancer. https://doi.org/10.3390/ijms26052049

The reason is the introduction of immunotherapy, that has revolutionized the treatment of NSCLC, from advanced stages to early stages, in combination with neoadjuvant and adjuvant therapies.[7] Specifically, the approval of immune checkpoint inhibitors (ICIs), targeting programmed cell death protein 1 (PD1) and programmed death-ligand 1 (PD-L1) such as Pembrolizumab and Nivolumab (anti-PD1) and Durvalumab (anti-PD-L1), has dramatically changed the treatment landscape for patients with NSCLC. Numerous studies in the metastatic setting have found higher response rates and more favorable survival outcomes when patients were treated with ICI monotherapy or ICI in combination with chemotherapy as compared with conventional chemotherapy alone. Despite these promising improvements, response to ICI treatment is not guaranteed in all patients, highlighting the need for predictive biomarkers.[6,7]

Numerous studies in the metastatic setting have found higher response rates and more favorable survival outcomes when patients were treated with ICI monotherapy or ICI in combination with chemotherapy as compared with

conventional chemotherapy alone. Despite these promising improvements, response to ICI treatment is not guaranteed in all patients, since the benefit is still limited to a subgroup of patients and, notably, often for a limited time: over 50% experience disease progression within the first 12 months of treatment ("early progressors"), and more than 68% do not reach five years of survival. Therefore, from the early years of ICI introduction, efforts have been made to identify biomarkers capable of predicting response and treatment benefit. To date, the only clinically validated biomarker is PD-L1 expression in tumor tissue, that is, the tumor proportion score (TPS). Tumors with TPS scores of 50% or above seem to have a higher likelihood of responding to ICI treatment alone, whereas those with a lower TPS percentage could best be treated with a combination of an ICI and chemotherapy.[8]

In recent years, liquid biopsy has shown numerous promising results, especially as many circulating biomarkers have shown a significant correlation with patient prognosis in several retrospective studies (e.g., searching for biomarkers such as immune checkpoints like PD-L1, B7-H4, B7-H3), but with still discordant results. Liquid biopsy can help identify mutations, but its sensitivity and specificity are still significantly lower compared to what can be obtained from tissue sampling and the accuracy further reduces when the disease is limited to the chest (stage I, II, or III). Moreover, although it seems like the ideal tool for evaluating systemic disease, predicting the response of an individual lesion to treatment is not currently possible.[8] A substantial aid can be provided by radiomics, the emerging discipline that focuses on the quantitative analysis of digital images, more specifically extracting features from radiological images to obtain information on clinical outcomes, such as prognostic and predictors factors of treatment response.[9]

Radiomics has emerged as a field of research that involves the process of converting standard-of-care images into quantitative image-based features that can be merged with other data sources and subsequently analyzed using conventional statistical or artificial intelligence (AI) methods. Radiomics is based on the hypothesis that digital features correspond to microscopic characteristics of the tumor. In radiomics, a pixel (in 2D images) or voxel (in 3D scans) refers to a single data point that represents specific information (e.g: intensity value, texture…) within the radiological images.[10] Radiomics extracts quantitative information by analyzing the relationships between these pixels, looking at patterns of signal intensity and texture. *(Figure 5)* It is, therefore, a non-invasive approach that allows the analysis of tumor characteristics, is repeatable over time, and permits the analysis of the lesion in its entirety, both its phenotype and its tumor microenvironment (unlike the small tissue samples obtained from biopsies).[8,10] As radiomic features capture biological and pathophysiological information, they have been shown to support rapid, accurate and noninvasive biomarkers for lung cancer risk, diagnosis and histology prediction, prognosis, and treatment response planning and monitoring.[11] The first step in radiomics is the segmentation, either manual or semi-automatic, of the lesion, meaning selecting the volume of interest (VOI), from which radiomic features are subsequently extracted. Given the enormous amount of data obtained from feature extraction, the use of artificial intelligence with “supervised” (machine learning) and “unsupervised” (deep learning) algorithms can significantly improve and speed up the analysis of these features.[11] Radiomics faces challenges due to variability in imaging acquisition, reconstruction, segmentation, software, and post-processing.[12] Standardization efforts like the Image Biomarker Standardization Initiative, quality control, and software harmonization aim to

reduce inter-study variability. Assessing segmentation impact is crucial, with the radiomics quality score by Lambin *et al.* and large language models like METhodological RadiomICs Score (METRICS) providing quality assessments.[33]

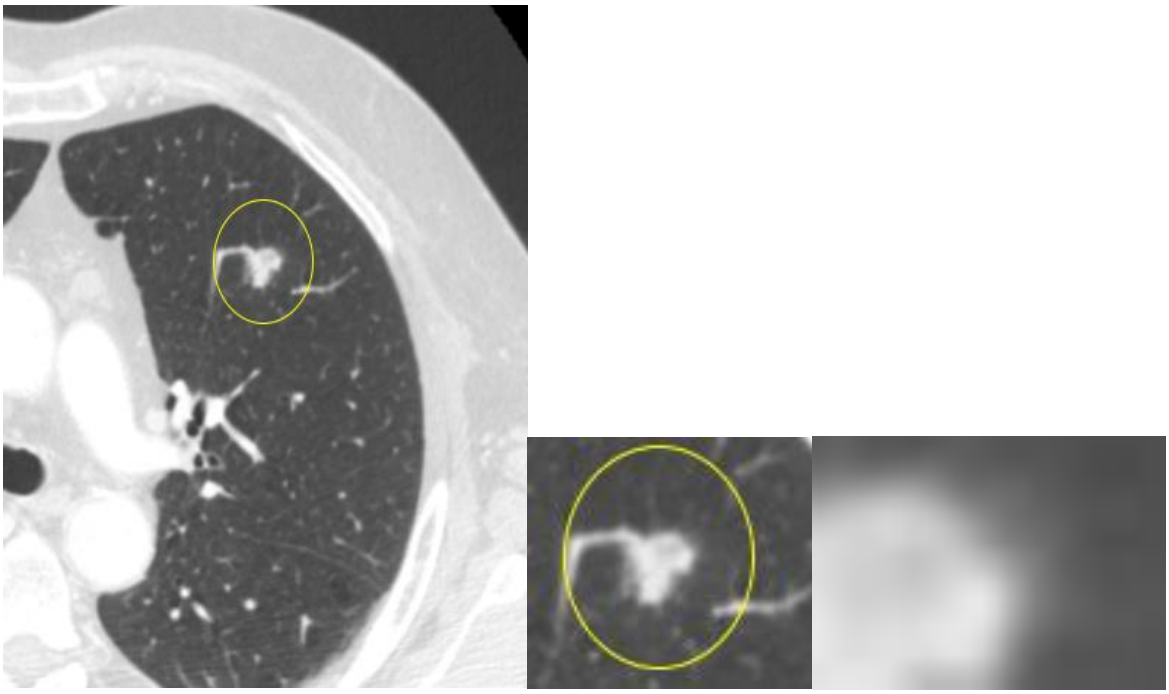


Figure 5: An example of lung nodule in left upper lobe and the magnification to pixel: In radiomics, a pixel (in 2D images) or voxel (in 3D scans) refers to a single data point that represents specific information related to the microscopic characteristics of the tumor (e.g: intensity value, texture).

More recently, delta-radiomics, a sub-specialty of radiomics, has been gaining increasing interest in clinical practice: delta-radiomics involves the analysis of features from the same region of interest at different timepoints to evaluate their variations (Δ) over time. Unlike radiomics, which is a "static" analysis, delta-radiomics studies phenotypic changes in tissue that may occur, for example, with the introduction of a new line of therapy, and this approach is particularly useful during immunotherapy, where the kinetics of tumor dimensional changes are not always associated with treatment response. In several studies, combining delta features with baseline features enhances

predictive accuracy for outcomes such as overall survival or treatment response.[13]

Since immune checkpoint inhibitors (ICIs) were introduced into clinical practice, significant efforts have been made to identify predictive biomarkers of response. To date, the tumor proportion score (TPS) of programmed death-ligand 1 (PD-L1) protein expression remains the only validated biomarker for guiding the treatment of oncogene-wild type non-small cell lung cancer (NSCLC). However, the diagnostic reliability of this marker has been frequently questioned due to its spatial (intra- and inter-lesions) and temporal variability. Indeed, the response prediction accuracy of PD-L1, measured by the area under the curve (AUC), is approximately 0.65.[14] In recent years, numerous studies have explored alternative biomarkers that may offer improved accuracy and reduced invasiveness. Despite these efforts, no additional biomarkers have been approved for routine clinical use yet. The ability to predict treatment response at the level of individual lesions could open new paths in the management of both advanced and early-stage NSCLC. This approach would enhance treatment personalization for each patient, potentially down to the level of single tumor lesions. However, due to the technical and anatomical challenges associated with obtaining adequate tissue samples in patients with NSCLC, the search for less invasive methods capable of assessing the entire disease—lesion by lesion—has gained increasing interest. In this context, radiomics may offer a promising solution. In computer tomography (CT) scan, the features extracted, often imperceptible to the human eye, reflect the interaction between X-rays and the body's tissues and cellular structures.[11,15]

The aim of this study is to assess whether the combination of radiomic features extracted from CT scans and artificial intelligence algorithms can

predict response to immunotherapy on a per-lesion basis. This approach differs from previous studies, which primarily focused on evaluating the overall prognosis of the patient based on radiomic analysis. The underlying hypothesis is that each lesion may exhibit distinct characteristics—such as variations in vascularization, lymphocytic infiltration, and stromal composition—that set it apart from others. This heterogeneity implies that the response of only one lesion cannot reliably inform the response of the disease as a whole, whether in the same organ or in different ones, and is even less indicative of overall patient survival.

2. Materials and Methods

2.1. Patients

Enrolled patients must be aged ≥ 18 years, and the selected corresponding lesions must meet the following criteria:

- Histological diagnosis of non-small cell lung cancer (NSCLC);
- PD-L1 expression equal or greater than 50% (TPS ≥50%);
- Absence of oncogenic driver mutations such as EGFR, ALK, MET, RET, and HER2;
- Baseline CT scan available;
- Presence of at least one measurable tumor lesion according to RECIST version 1.1 (Response Evaluation Criteria in Solid Tumors).

The present retrospective study included adult patients with advanced NSCLC who had been treated with an antiPD1 monotherapy as first or second line at IRCCS San Martino Polyclinic Hospital in Genoa, Italy (*main cohort-main*

center). A baseline contrast-enhanced CT scan (chest-abdomen-pelvis protocol) performed within 30 days of the starting anti-PD1 therapy was required for inclusion in the study. Therefore, patients with baseline examinations such as non-contrast CT scans or PET-CT are excluded.

Each patient provided written informed consent for the collection of clinical data and the transmission of radiological images for study purposes.

The study included 125 patients enrolled across three different centers in Italy: 80 at main center, the IRCCS San Martino Polyclinic Hospital in Genoa (training set), 23 at the University Hospital of Parma (TestSet1), and 22 at the Papardo Hospital in Messina (TestSet2). *(Table 1).*

In training set, a total of 101 lung lesions and 61 lymph node sites were analyzed, that have been segmented by three radiologists (E.B., with 10 years of experience, S.P. with 7 years of experience, E.N. with 3 years of experience).

The testing external cohorts included 38 lung lesions and 23 lymph nodes in the TestSet1, and 26 lung lesions in the TestSet2 *(Table1).*

Overall, 83 patients were treated with pembrolizumab and 42 with nivolumab. Based on treatment outcomes, 66 patients were labelled as progressors (P), while 59 were labelled as non-progressors (NP).

Up to a maximum of two lesions per patient were included (for example, the primary lesion and a pathological lymph node), to increase the sample size while minimizing the risk of *overfitting*, due to biological similarity between lesions from the same individual. *Overfitting* could be a bias in the results since we presume that lesions in the same patients can share biological similarity and reasonably can also share similar response to treatment *(Figure 1)*.

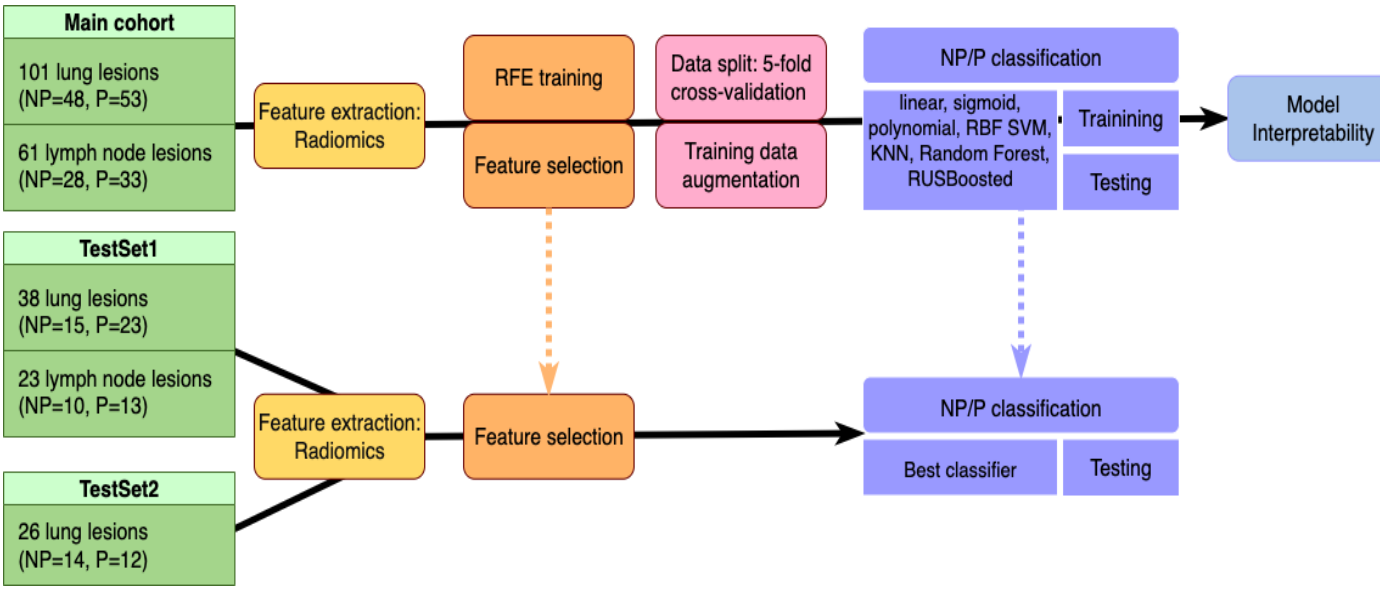


Figure 1: study pipeline and description.

Patients' characteristics		Main cohort 80 pts N (%)	Testset1 23 pts N (%)	Testset2 22 pts N (%)
Gender	M	60 (75)	14 (61)	13 (59)
	F	20 (25)	9 (39)	9 (41)
Mean Age		69 years	70 years	70 years
Smoking status	Smoker	23 (28)	10 (44)	9 (41)
	Former	54 (68)	9 (39)	9 (41)
	Never	-	4 (17)	4 (18)
	Unknown	3	-	-
Histology	ACK	53 (65)	15 (65)	14 (64)

	SCC	19 (23)	7 (31)	4 (18)
	Other	8 (10)	1 (4)	4 (18)
PD-L1	>50%	48 (60)	12 (52)	9 (41)
	<50%	1 (1)	10 (44)	5 (23)
	Unknown	31 (38)	1 (4)	8 (36)
Anti-PD1 drug	Pembro	49 (60)	23 (100)	11 (50)
	Nivo	31 (38)		11 (50)

Table 1. Patients' main characteristics (AC=adenocarcinoma; SCC= squamous cell carcinoma; Pembro=pembrolizumab; Nivo = nivolumab)

During the first half of the second year of the PhD program, the main task consisted of enriching the existing training-set database required to train the model for the development of the artificial intelligence (AI) algorithm. Indeed, by the end of the first PhD year, one of the major critical issues that emerged was realizing that the training set cohort was too "homogeneous," as the selected and segmented target lesions predominantly exhibited a non-progression response pattern. To optimally train the AI algorithm, it is essential to build a dataset that is as large as possible while also reflecting, as accurately as possible, the "real world scenarios" of the disease in the clinical practice.

For this reason, the database was enriched with a retrospective cohort of 66 patients who exhibited a short "time to treatment failure" (TTF), namely those patients who prematurely discontinued immunotherapy due to confirmed disease progression within 12 months of treatment initiation (early progressors).

2.2. Evaluation of Response

Each lesion was segmented and assessed individually. The first clinical challenge is precisely defining the clinical response, as there are a variety of response patterns during immunotherapy, making it difficult to rely solely on volumetric changes of the lesion. For example, an initial increase in lesion volume could be due to the recruitment of inflammatory cells in the tumor tissue and therefore does not necessarily correspond to disease progression (called "*pseudoprogression*").

It is important to note that each target lesion was segmented and assessed individually: at the beginning of the project, in accordance with RECIST 1.1 criteria, a lesion was considered a “responder” if it shows a reduction of more than 30% in its maximum diameters and no subsequent progression. Conversely, a lesion was considered resistant or a “non-responder” if it shows an increase of more than 20% in its maximum diameters. Lesions classified as “stable disease” therefore include a heterogeneous group, with dimensional changes ranging from a reduction of less than 30% to an increase of less than 20%. (*Figure 1)*

During the second year of the PhD program, it was considered appropriate to modify the parameters of the above classification to improve the accuracy of the predictive model, as it became evident in clinical practice that the “stable disease” group often tends, over time, more toward progression than non-progression—especially after a prolonged period of disease stability.

It was quickly realized that the RECIST (Response Evaluation Criteria in Solid Tumors) criteria were insufficient to explain the progression of the disease.

To dichotomize lesion-level responses to immunotherapy, we defined a lesion as “non-progressor” if it showed a minimum reduction greater than 10% at

the first timepoint with no subsequent progression. Conversely, lesions exhibiting an increase in their major diameter greater than 10% at the first timepoint were considered as “progressor”. Lesions with stable major diameters were labelled based on their evolution over the following six months: those showing either a subsequent reduction or non-progression were considered “stable disease” and included in the “non-progressor” group, while those with a dimensional increase were included in the “progressor” group.

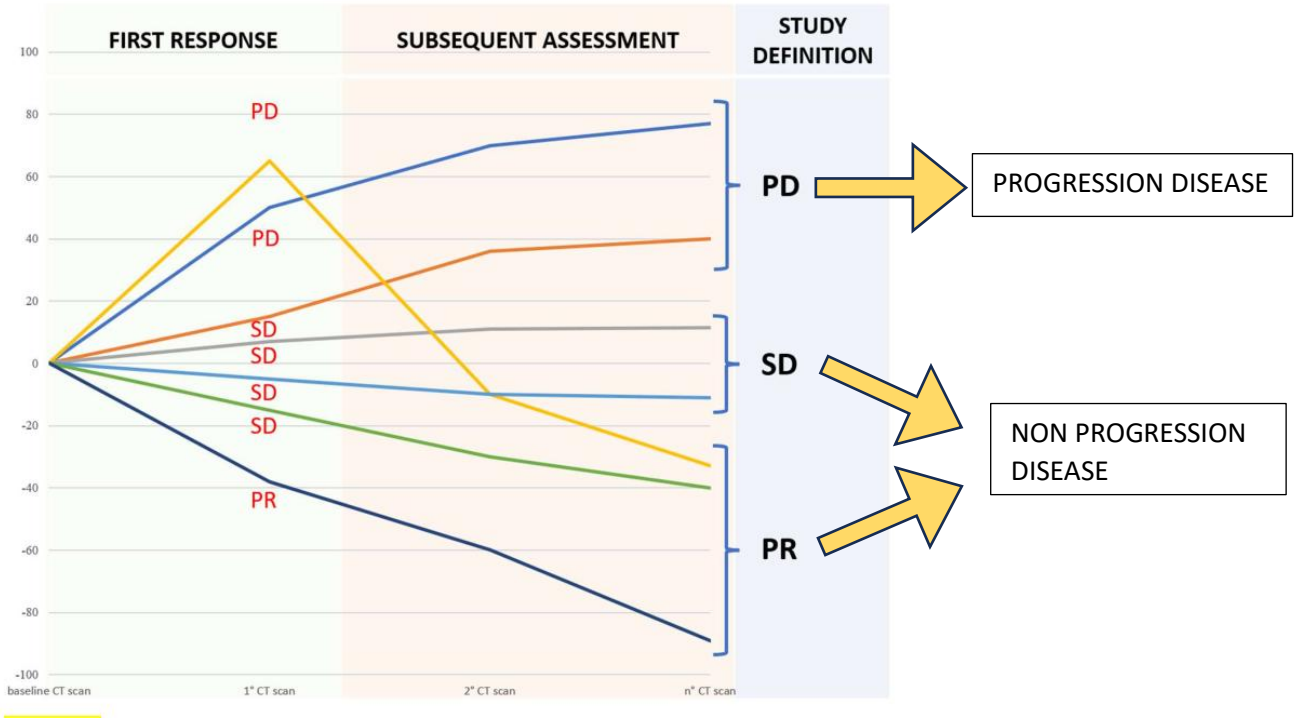


Figure 1.: Immunotherapy response simplified into a binary system: PD (Progressive Disease) vs. non-PD (Stable Disease + Partial Response).

2.3. External Validation Set

The best predictive model was tested on two independent external datasets, referred to as TestSet1 and TestSet2, composed of patients selected according to inclusion criteria used for the main cohort. TestSet1 was composed of 38 pulmonary lesions from patients treated and enrolled at the University

Hospital of Parma, Italy. For this dataset, CT scan selection, reformat, segmentation and radiomic analysis were performed on-site by two experienced thoracic radiologists (GM and MB). TestSet2 includes 26 pulmonary lesions from patients treated and enrolled at the Papardo Hospital in Messina, Italy. CT images reformat and segmentation of these lesions were then performed at IRCCS San Martino Polyclinic Hospital in Genoa by the same radiologists who analyzed the main cohort. Data from both TestSet1 and TestSet2 were normalized using the same procedure applied to the main cohort to ensure consistency across datasets.

2.4. CT protocol and images reformatting

All patients included in the main cohort underwent a baseline contrast-enhanced CT scan, within 30 days prior to the initiation of the treatment with immunotherapy. The baseline imaging consisted of a contrast-enhanced total-body CT scan (including chest-abdomen-pelvis and brain). Patients whose baseline imaging consisted of non-contrast CT scans or PET-CT are therefore excluded from the study. CT acquisition was performed on two 16-rows and a 64-rows scanner (LightSpeed and Optima, Ge Healthcare, Chicago, Illinois) after contrast media injection (80-120 ml, Iopamiro 370, Bracco) during the portal-venous phase (approximately 70–80 seconds after intravenous contrast injection). Acquisition parameters were: tube voltage 120 kV, slice thickness 1.25 mm, FOV 35-50 cm, matrix 512×512. Tube current automatic modulation was employed. DICOM images were exported on a dedicated offline workstation (AW Server 4.6, GE Healthcare, Chicago, Illinois and Syngo.via workstation, Erlangen, Siemens Helthineers) and reformatted on the axial plane at 3.0 mm of slice thickness and FOV 32 cm.

CT examinations of the TestSet1 from Papardo Hospital in Messina were performed using 128- and 64-detector row scanners (Optima 660, GE Healthcare, Chicago, IL, USA) following intravenous administration of contrast medium (70–130 mL of Iomeron 400, Bracco; Ultravist 370, Bayer; or Xenetix 350, Guerbet). Acquisition parameters were as follows: tube voltage, 120 kV; slice thickness, 2.5 mm; field of view, 35–50 cm; and matrix size, 512 × 512. Automatic tube current modulation was applied.

CT examinations of the TestSet2 from University Hospital of Parma included 23 patients, 2/23 were evaluated on a non-contrast enhanced CT scan. The other 21/23 patients underwent contrast enhanced CT scan, including these phase acquisitions: 4/21 had also availability of an arterial phase, while the other 17/21 were scanned just in a portal-venous phase.

All images, from this moment on, have been anonymized and coded with a progressive numeric code preceded by the acronym of the center of origin. Each lesion has been coded as lung or lymph node followed by a progressive number. For each center there is a decoding file of the patient codes for any additional research.

2.5. Segmentation and Radiomic analysis

Reformatted images were transferred on an off-line workstation. Measurable target lesions were selected (measured by maximum axial diameters according to RECIST 1.1 criteria), typically including the primary tumor, pathological lymph nodes, and, more rarely, hepatic metastases (after excluded for the limited data).

Segmentation and features extraction were carried out by the same software in the different centers: segmentation was performed using 3D Slicer 4.10.1

(http://www.slicer.org) and radiomic features were extracted with the Pyradiomics extension. (*Figure 1*)

A radiologist (EB) with 10 years of experience in thoracic imaging (6 years of experience in radiomics) and two radiologists with respectively 7 (SP) and 3 (EN) years of experience in thoracic imaging blinded to clinical data, delineated a Volume-of-Interest (VOI) around each lesion using a semi-automated method and a slice-per-slice approach. *(Figure 2)*

Hyperdense structures, such as bone or metal objects, were excluded from the segmentation; if present, macroscopic air was included within the VOI, even in case of cavitation at the consequent CT, but major lobar branches were excluded. Some of these lesions are more challenging to segment due to the presence of adjacent vascular or bronchial structures, or because of their proximity to the chest wall or mediastinum. An even greater challenge is represented by the distinction between the neoplastic lesion and the non-pathological adjacent components of pulmonary consolidation, for instance in cases where the lesion causes lung atelectasis or post-obstructive pneumonia. In such cases, the key role of the radiologist is to accurately discriminate among the different components in order to avoid overestimating the disease volume. (*Figure 3)*

A total of Eight-hundred fifty-one radiomic features were extracted for each lesion. The mathematical definitions of the adopted features are available for consultation on Pyradiomics' website (https://pyradiomics.readthedocs.io/en/latest/features.html). The features reported patient by patient (completely anonymized) are reported in the database attached to the supplementary materials of this paper.

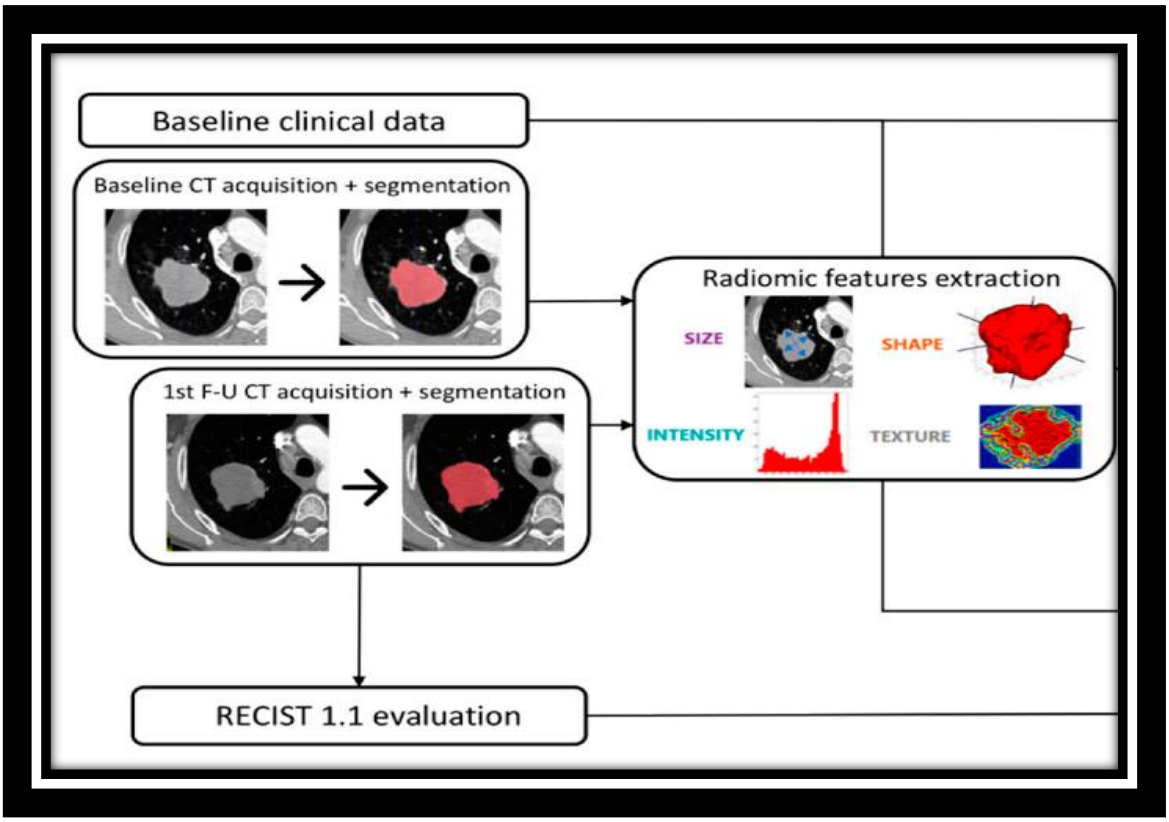


Figure 1: Workflow summarizing the main steps: baseline CT evaluation with lesion measurement according to RECIST 1.1, lesion segmentation using 3D-Slicer (Pyradiomics), and subsequent radiomic features extraction.

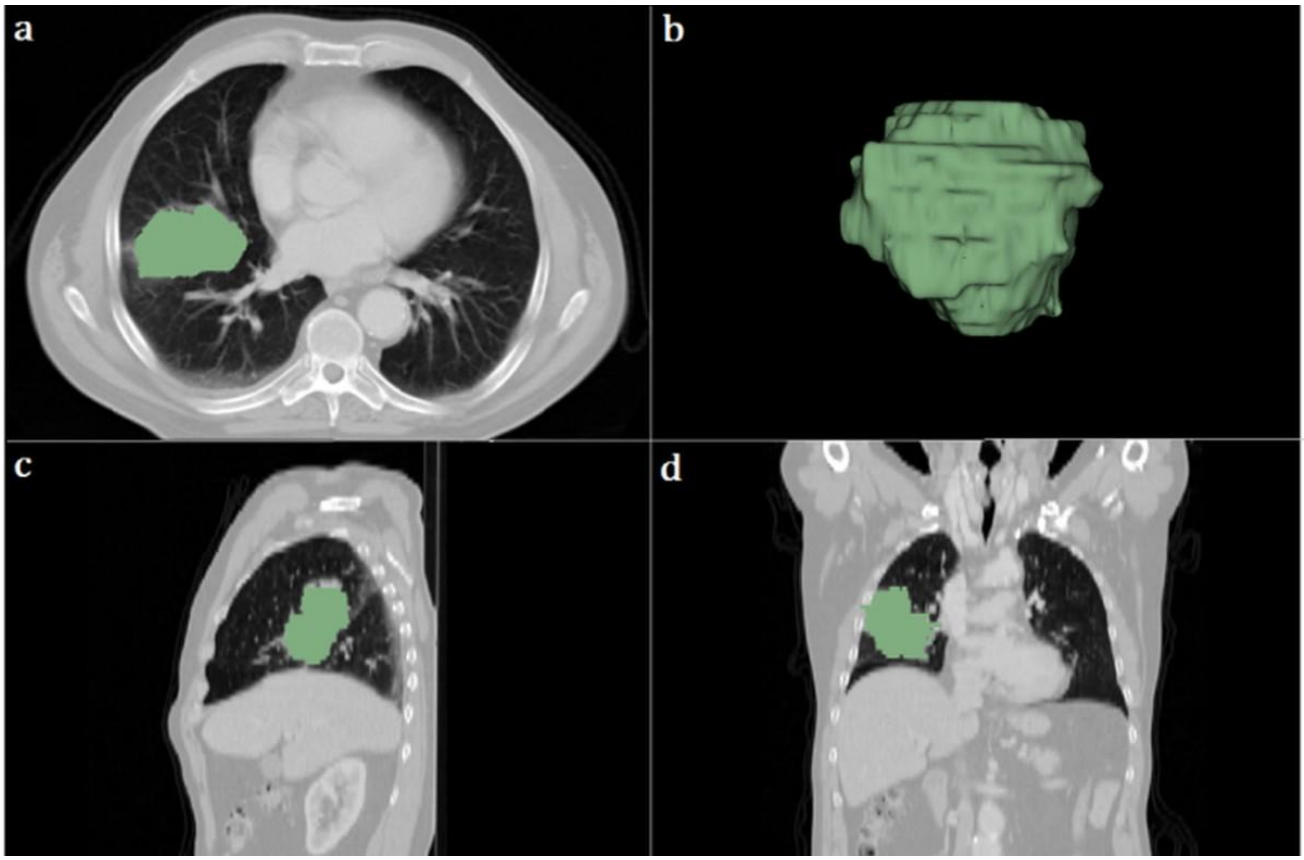


Figure 2: an example of lung mass segmentation using 3D-Slicer, with multiplanar views and 3D reconstruction.

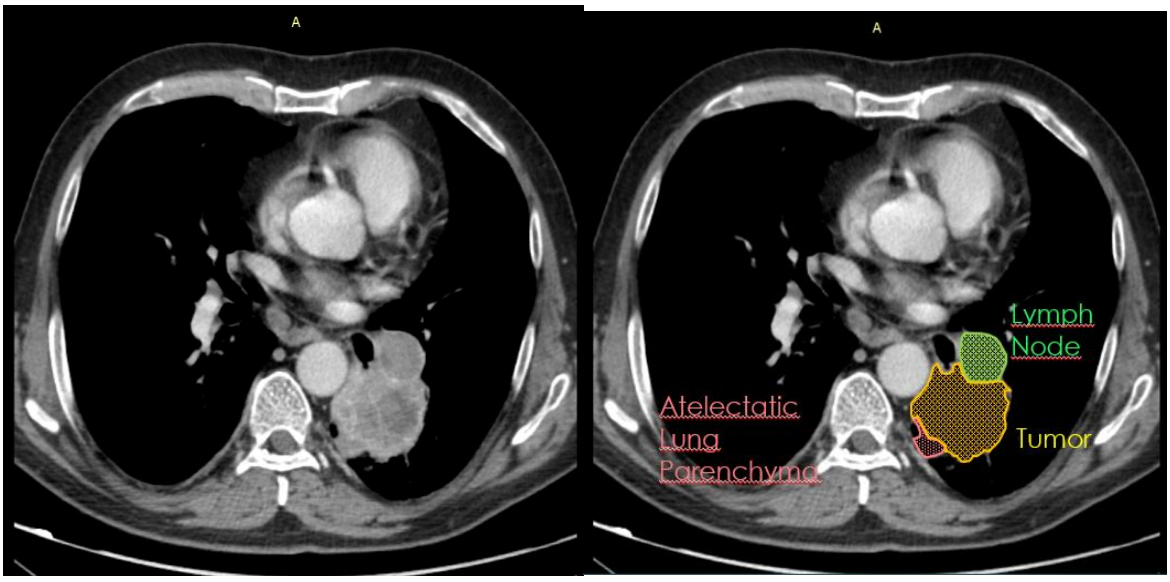


Figure 3: Baseline CT of a left lower parahilar lung mass showing a heterogeneous mass with three distinct components that highlight the difficulty of segmentation: primary lung tumor (yellow), pathological hilar lymphadenopathy (green), and atelectatic lung parenchyma (red).

2.6. Feature selection

During the second year of the PhD program, with the support of the Laboratory of Computational and Statistical Methods at the University of Genoa (Dr. Rosanna Turrisi), the feature selection and analysis have been conducted. In the first step, the radiomic features were standardized using Z-score normalization for all lesions.

Then three different approaches were investigated for feature analysis and test set creation, to eliminate redundant information and to select the more relevant features, specifically the most predictive of the immunotherapy treatment response.

The first approach involved the analysis of all extracted features—whose number greatly exceeded the sample size—without any prior selection, using various established machine learning algorithms.

Subsequently, the second approach included a feature classification with the use of Recursive Feature Elimination (RFE), a widely used method for reducing numerosity by identifying the most informative features for a given classification task. The RFE process made use of a Linear Support Vector Machine (SVM) algorithm to classify lesions into non progressor and progressor simultaneously.[15] SVM was selected because of its well-documented effectiveness in high-dimensional settings, such as radiomic feature analysis. The RFE process works by first ranking features based on their contribution to classification. In each iteration, the least informative features are eliminated, and the process is repeated with the remaining features until an optimal subset of predictive features is identified. By iterating this process, the algorithm gradually converges toward the most predictive features.[15,16]

Highly correlated and redundant features were removed, and those specifically relevant to the clinical objective were retained. Features with a correlation greater than 90% were excluded, resulting in the selection of n = 362 features for “lung” lesions, n = 361 features for “node” lesions, and n = 180 features for the combined “lung + node” dataset. *(Table 1 and 2)*

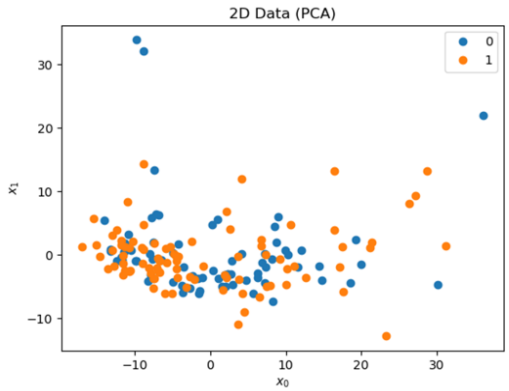
The third approach combined feature selection and RFE with the use of multiple classifiers, such as linear SVM, SVM with Non-linear Kernels (e.g., Radial Basis Function (RBF), Sigmoid), Random Forest, RUSBoosted Trees (RUSBoost). Using multiple classifiers in the RFE process allows for:

- Validation of feature relevance across different models;
- Improved generalization to unseen data;
- Reduced model bias associated with a specific classifier;
- Assessment of stability of selected features under different decision boundaries.

This multi-model approach is especially important in biomedical and radiomic contexts, where models must be both accurate and robust across heterogeneous patient populations and imaging protocols.[17]

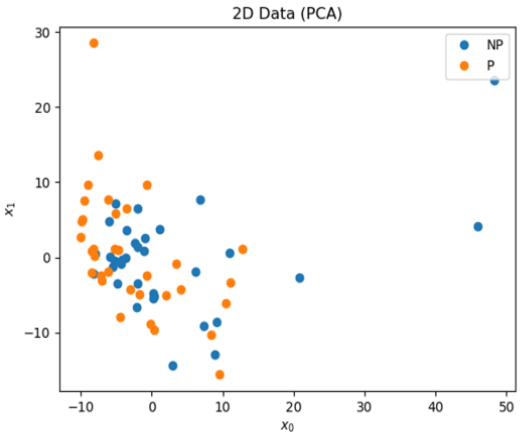
Moreover, the same analytical methodology was applied both to the separate cohorts of pulmonary (“lung”) and lymph node (“node”) lesions, as well as to the combined dataset (“lung + node”).

LUNG DATASET



- 4 features costanti sono state eliminate:
('wavelet-HHH', 'firstorder', 'Entropy'),
('wavelet-HHH', 'glcm', 'JointEntropy'),
('wavelet-HHH', 'glcm', 'SumSquares'),
('wavelet-HHL', 'glcm', 'DifferenceEntropy')

NODE DATASET



- 2 campioni senza label sono stati esclusi
- 7 features costanti sono state eliminate:
('wavelet-HHH', 'firstorder', 'Entropy'),
('wavelet-HHH', 'glcm', 'JointEntropy'),
('wavelet-HHH', 'glcm', 'SumEntropy'),
('wavelet-HHL', 'firstorder', 'Entropy'),
('wavelet-HHL', 'glcm', 'JointEntropy'),
('wavelet-HHL', 'glcm', 'SumEntropy'),
('wavelet-LHL', 'glcm', 'DifferenceEntropy')

Table 1,2: Feature selection had the aim to eliminate highly correlated or redundant features and to select features specific to my objective.

2.7. Classification of Response to Immunotherapy Treatment

Predictive Model Development

In this study, we developed predictive models based on machine learning algorithms to classify lung and lymph node lesions as either progressive (P) or non-progressive (NP) during immunotherapy. The main goal was to anticipate,

based on radiomic features, which lesions would exhibit disease progression, and which would not.

To achieve this, several classification algorithms were tested *(Table 1 and 2)*, including:

- Support Vector Machines (SVM) with different kernels (linear, sigmoid, polynomial, and RBF – Radial Basis Function);
- k-Nearest Neighbors (k-NN);[18]
- Random Forest;[19]
- Random Under-Sampling Boosted (RUSBoosted), an algorithm that combines boosting with under-sampling of the majority class.[20]

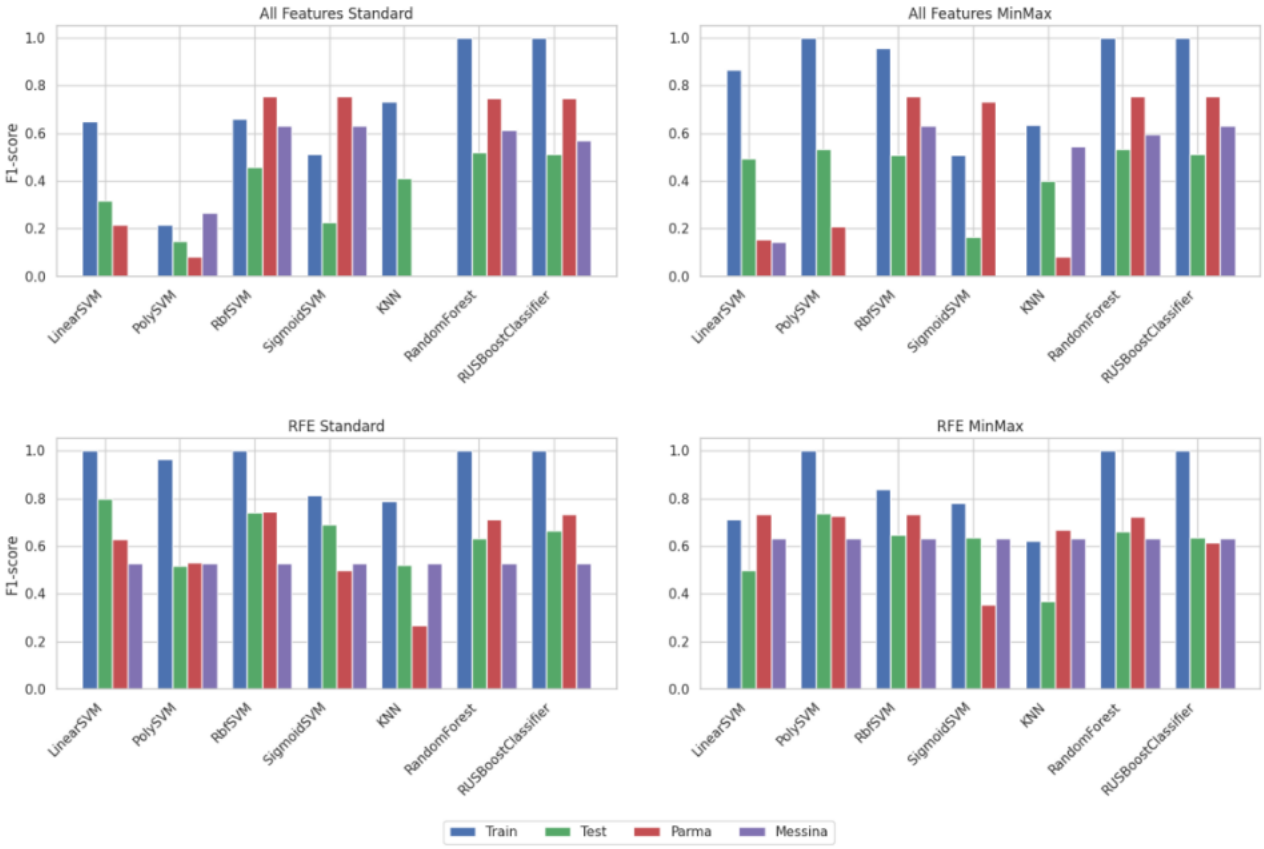


Table 1: comparison between different classifiers for lung lesions.

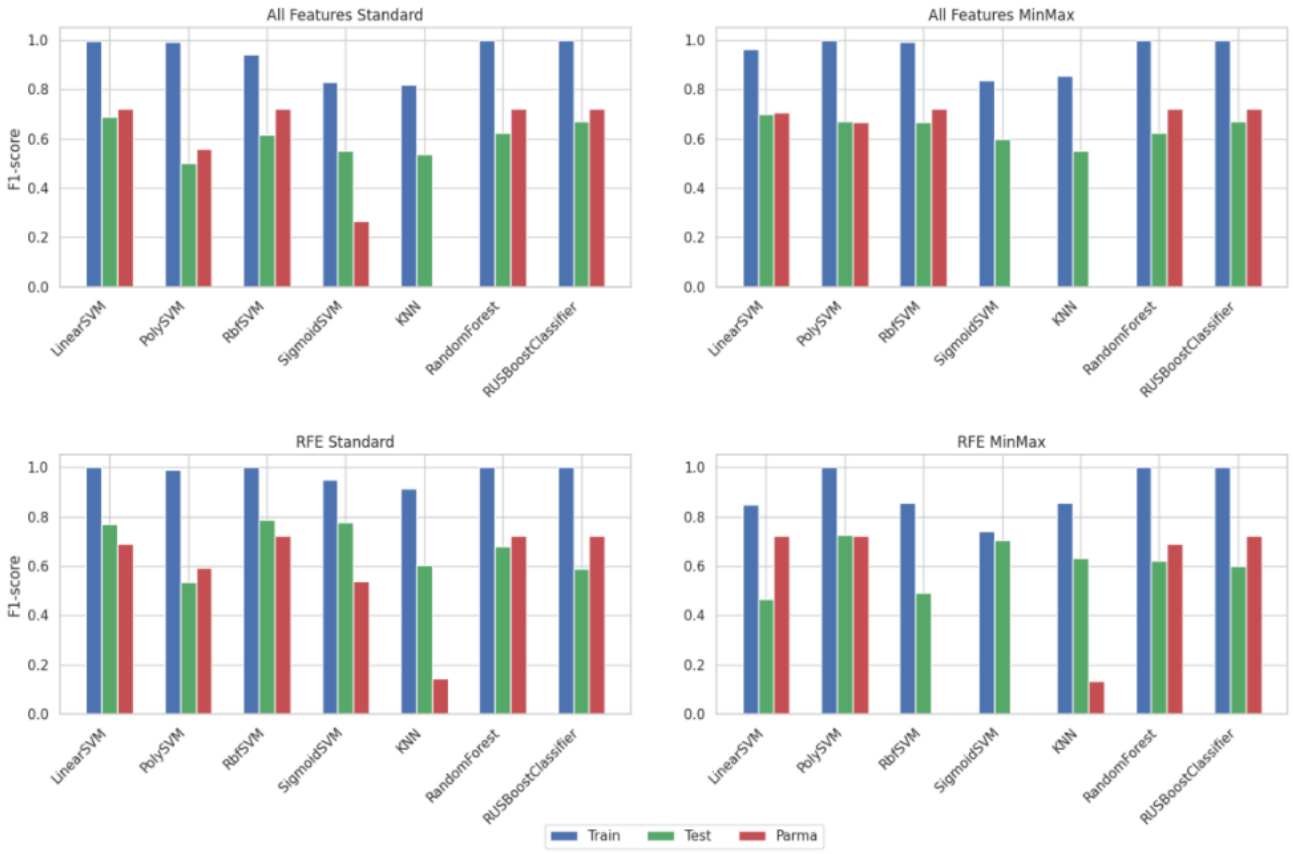


Table 2: comparison between different classifiers for lymph nodes.

The dataset was divided into training and testing subsets using 5-fold cross-validation (k=5, random state=0), a method that helps reduce the risk of overfitting. Overfitting occurs when the model has low bias but high variance, therefore performing well on training data but poorly on unseen data.

A common issue in clinical datasets is class imbalance. For example, in our case, the number of non-progressive lesions tends to be much higher than the progressive ones. To address this, we applied the Synthetic Minority Over-sampling Technique (SMOTE), which creates synthetic examples of the minority class to balance the dataset. SMOTE was used for all models except for RUSBoosted, which handles imbalance inherently through under-sampling.[21]

For each algorithm, hyperparameter tuning was performed using a grid search method with 5-fold cross-validation (random state=123), in order to identify the combination of parameters that yielded the best performance.

Evaluation metrics

To select the most effective model, F1-score has been chosen as the main evaluation metric. The F1-score is particularly appropriate for imbalanced datasets, as it balances precision (how many of the lesions predicted as progressive truly are) and recall (how many of the actual progressive lesions were correctly identified).

Once the best-performing model was identified based on the training set, we evaluated it on both the testing set and two external independent TestSets (referred to as TestSet1 and TestSet2) to assess its generalizability. Performance was assessed using the following metrics:

- Confusion Matrix;
- Precision;
- Recall;
- AUC (Area Under the Curve), which indicates the model's ability to distinguish between classes;
- AUPRC (Area Under the Precision-Recall Curve), which is particularly useful in imbalanced scenarios.

Model interpretability

Beyond model performance, it is essential— especially in clinical applications —to understand which features influence the model's decisions. A model must not only be as accurate as possible but also interpretable.

To this end, the SHapley Additive exPlanations (SHAP) method was adopted, because it allows to quantify the contribution of each feature to individual predictions. SHAP assigns an "importance value", known as a SHAP value, to each radiomic feature. This value reflects how much that feature has contributed (positively or negatively) to the model's output for a given instance.[22]

In addition, SHAP provides intuitive visualizations that help clinicians and researchers to better interpret the model's reasoning and the role of individual radiomic features in decision-making.[23]

3. Results

3.1. Model Performance on the Main Cohort

To assess model performance on the main cohort (training set), various machine learning algorithms were tested and evaluated. Among these, the Support Vector Machine (SVM) with a sigmoid kernel emerged as the best-performing classifier for lung lesions. This model achieved an F1-score of 0.77, indicating a balanced trade-off between precision and recall. Specifically, the model demonstrated a precision of 0.75, meaning that 75% of the predicted progressive cases were correct, and a recall of 0.81, indicating that it correctly identified 81% of all true progressive lesions.

In terms of probabilistic performance, the Area Under the Receiver Operating Characteristic Curve (AUC) was 0.89, and the Area Under the Precision-Recall Curve (AUPRC) reached 0.90, suggesting a strong overall ability to distinguish between progressive and non-progressive lesions even in the presence of class imbalance.

The confusion matrix for this classification task, based on 5-fold cross-validation, is reported in Figure 2. The table on the left shows the prediction performance of the algorithm within the main cohort for lung lesions: the model misclassified 14 non-progressive lesions as progressive (false positives). (*Figure 2, left*)

For lymph node lesions, the best performance was obtained using a linear SVM. This model reached an F1-score of 0.86, a precision of 0.94, and a recall of 0.82, indicating both high specificity and sensitivity. The corresponding AUC and AUPRC were 0.95 and 0.96, respectively, confirming excellent discriminative power. The confusion matrix shows only 2 false positive classifications, underlining the model's robustness (*Figure 2, right*).

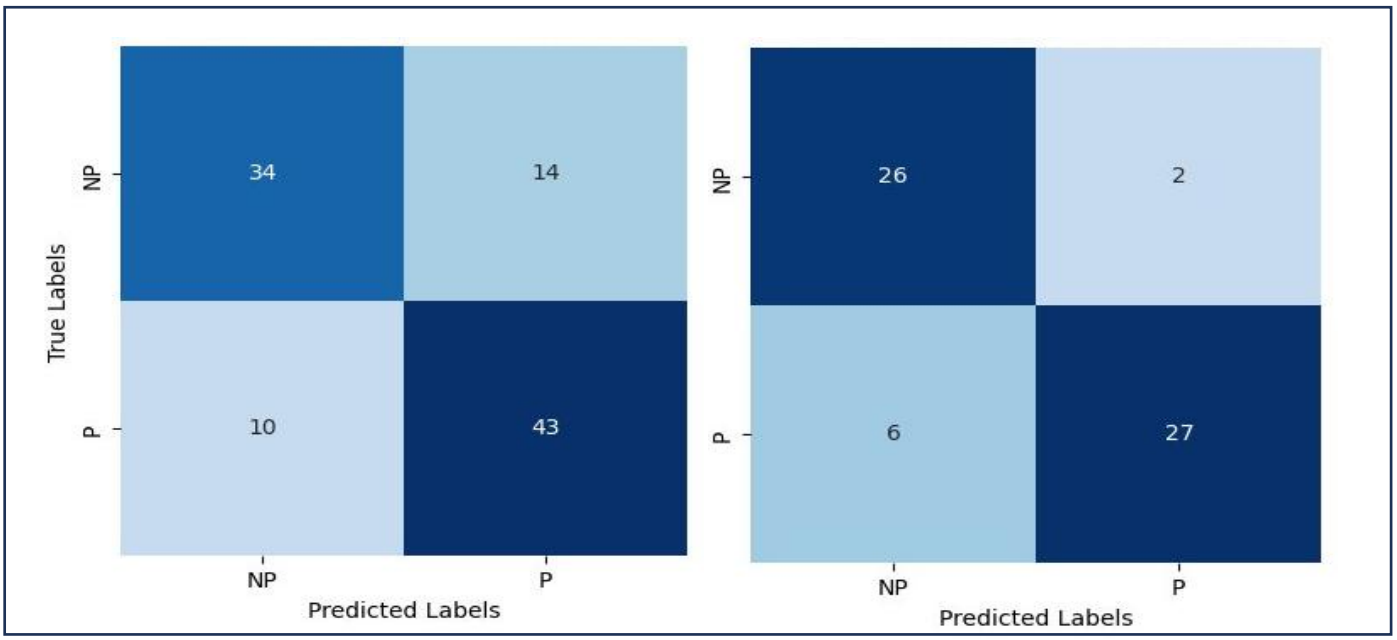


Figure 2.: confusion matrix reporting the prediction performance of the algorithm within the main cohort for lung lesions (right) and lymph node (left) lesions

3.2. External validation assessment

To test the generalizability and possibly standardization of the predictive models, we evaluated their performance on two external independent test sets.

For lung lesions, the best-performing model (SVM with sigmoid kernel) achieved a consistent F1-score of 0.69 on both TestSet1 and TestSet2. However, the precision remained relatively low at 0.59, while recall was high at 0.83, suggesting a tendency of the model to over-predict progression. The AUC values for TestSet1 and TestSet2 were 0.30 and 0.46, respectively, while AUPRC values were 0.41 and 0.58, indicating limited discrimination in the external data—possibly due to heterogeneity in CT-scan acquisition protocols or lesion intrinsic characteristics.

- On TestSet1, the model correctly classified 21 out of 38 lesions, with 13 false positives.
- On TestSet2, the model correctly identified 17 out of 26 lesions, with 7 false positives.

These results are visualized in *Figure 3*, which shows the confusion matrices for both external test sets. Despite moderate recall, the number of false positives remains considerable, highlighting limitations in model transferability.

For lymph node lesions, the prediction was higher with better results than the one for the lung lesions; the best model yielded an F1-score of 0.77 on TestSet1, with a precision of 0.62 and a perfect recall of 1.00—meaning that

all actual progressor lesions were correctly identified. The AUC and AUPRC were 0.60 and 0.81, respectively. According to the confusion matrix *(Figure 4*), out of 23 lesions, 15 were correctly classified, and 8 were false positives.

On TestSet1, for lung lesions, the model obtained:

- Precision = 0.59
- Recall = 0.83
- AUC = 0.30
- AUPRC = 0.41.

On TestSet2, for lung lesions, the model obtained:

- Precision = 0.59
- Recall = 0.83
- AUC = 0.46
- AUPRC = 0.58.

On TestSet1, for lymph nodes, the model obtained:

- Precision = 0.62
- Recall = 1.00
- AUC = 0.60
- AUPRC = 0.81.

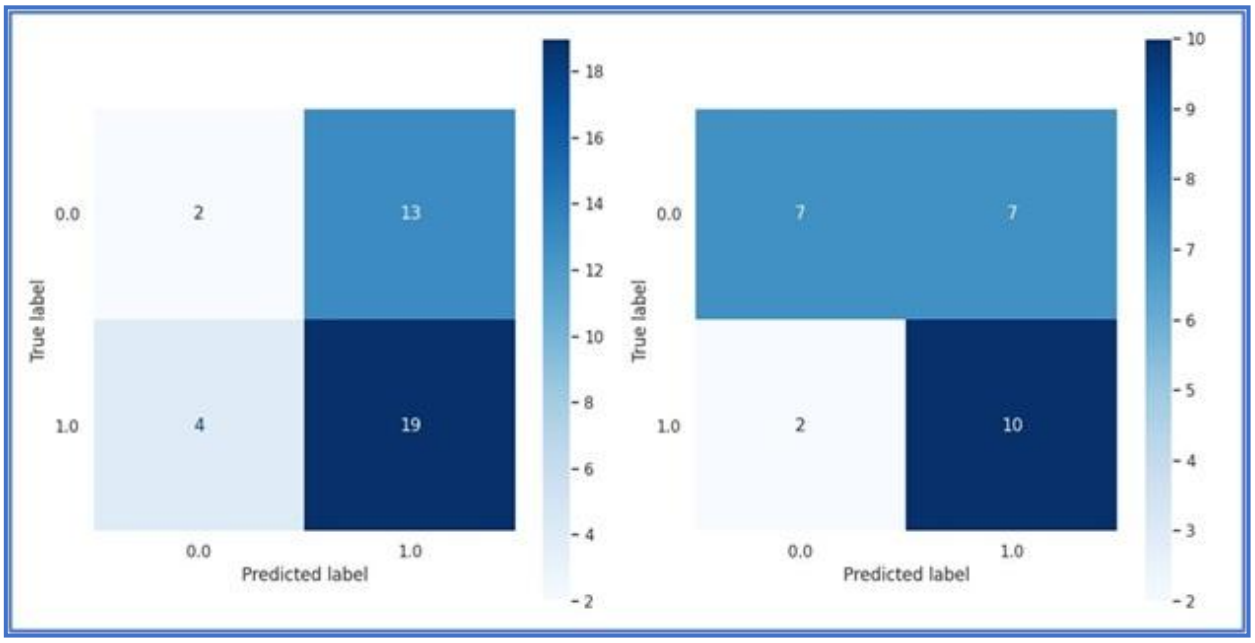


Figure 3: confusion matrix reporting the prediction performance of the algorithm based on lung lesions from TestSet1 (left) and TestSet2 (right).

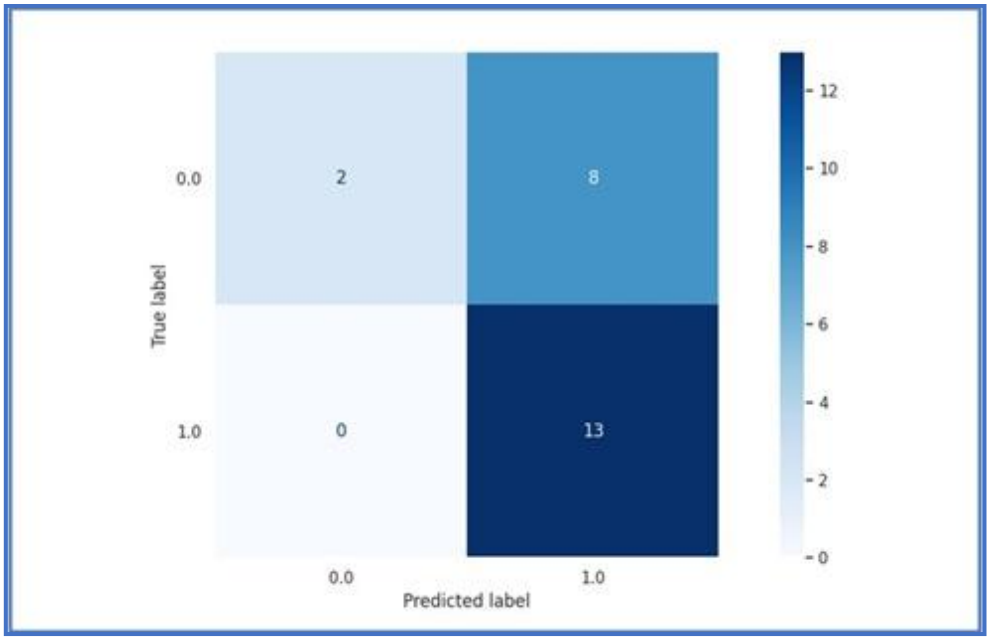


Figure 4 confusion matrix reporting the prediction performance of the algorithm on TestSet1 for node lesions.

Approach	*Testing set*	*TestSet1((Parma)*	*TestSet2(Messina)*
1	*71.2*	*73.1*	*51.6*
2	*66.7*	*63.4*	*63.8*
*3 SVM (*radial basis function - *RBF)*	*68.1*	*75.4*	*63.8*
3 SVM (sigmoid kernel)	*77.0*	*69.0*	*69.0*

Table 1: resume of F1score% using the three different approaches for lung lesions: SVM with sigmoid kernel showed the best results both in testing set and external validation TestSets.

Approach	Testing set	TestSet1 (Parma)
1	76.2	72.7
2	78.6	62.5
3 SVM (radial basis function - RBF)	70.1	72.7
3 SVM (sigmoid kernel)	86.2	77.6

Table 2: resume of F1score% using the three different approaches for lymph nodes: SVM with sigmoid kernel showed the best results both in testing set and external validation TestSet.

3.3. Model interpretation

Figures 5 and 6 depict the SHAP (Shapley Additive exPlanations) summary plots derived from lung lesions and lymph node lesions, respectively, within the main cohort. These visualizations highlight the relative contribution of the ten most predictive radiomic features to the model's decision-making process, by summarizing the distribution of their SHAP values across all samples.

In both plots, features are ranked along the y-axis in descending order of overall importance, with the most influential predictors appearing at the top. Each dot corresponds to an individual lesion, and its position on the x-axis indicates the SHAP value, which quantifies the magnitude and direction of the feature's effect on the predicted class.

For lung lesions (Figure 5), features such as *wavelet-LLH, first-order, Total Energy* emerged as particularly discriminative. High values of this feature (represented by red dots on the left side of the plot) increased the probability of classification as a "progressor." Conversely, lower values (depicted in blue dots) shifted the prediction toward the "non-progressor" class.

In lymph node lesions (Figure 6), the most predictive features included *wavelet-LHL, ngtdtm, Busyness*. Here, low values were strongly associated with the "progressor" class, suggesting that a reduced degree of intensity variation across the lesion's texture was indicative of disease progression.

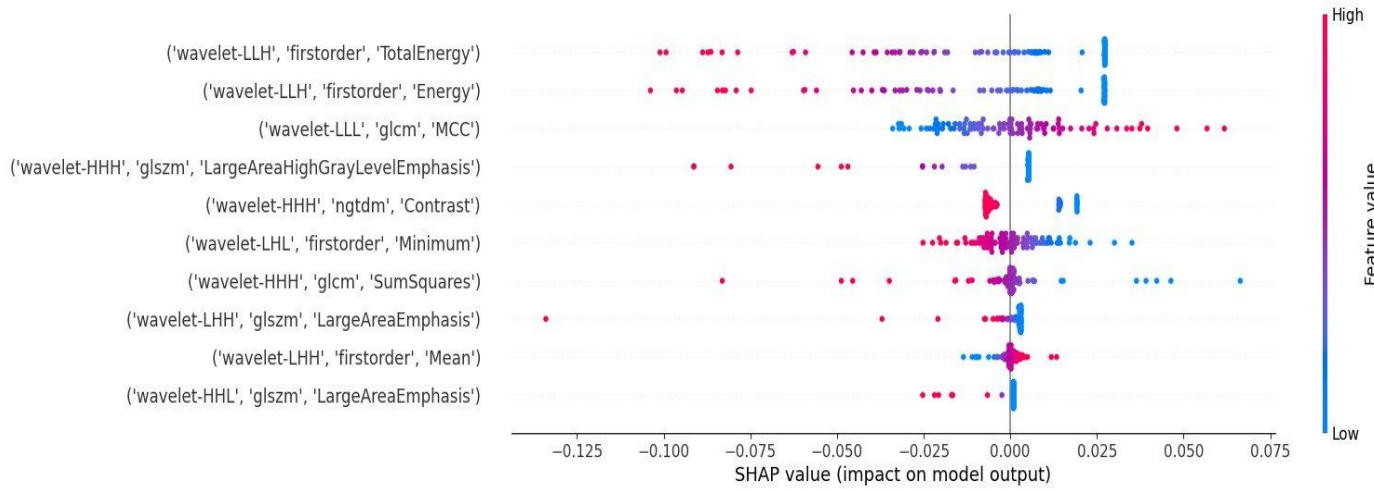


Figure 5. SHapley Additive exPlanations summary of the ten most predictive radiomic features for lung lesions.

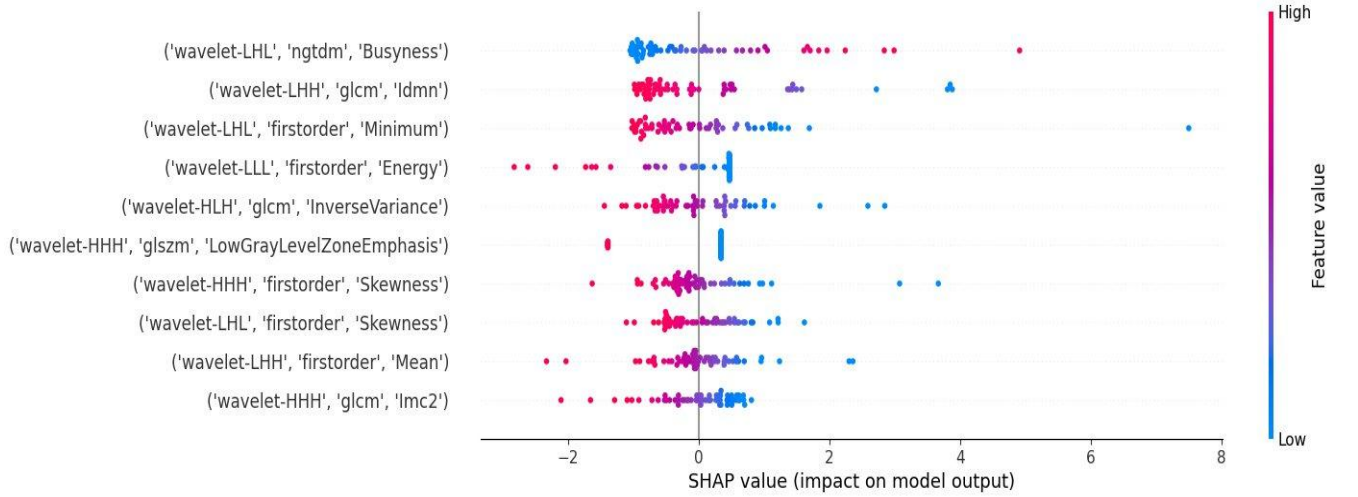


Figure 6. SHapley Additive exPlanations summary of the ten most predictive radiomic features for lymph nodes.

3.4. Predictive Features Interpretation

According to the SHAP-based features analysis, non-progressor and progressor lung lesions presented two distinct radiomic phenotypes**.**

Progressor lung lesions were characterized by elevated values of *Total Energy*, *Energy*, and *Large Area High Gray Level Emphasis*. These findings suggest a

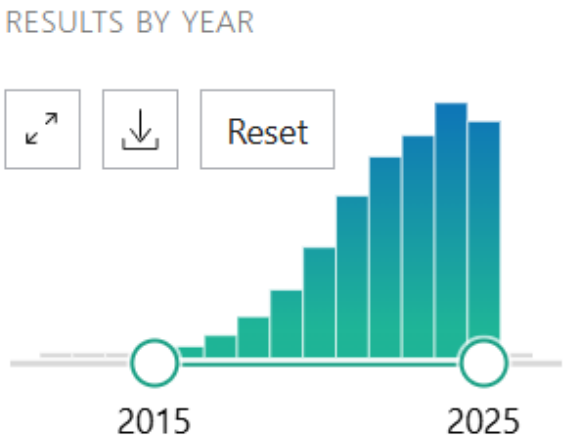
high degree of intralesional heterogeneity, marked by localized regions of high attenuation. Additionally, the increased *Contrast* values indicate pronounced intensity variations, consistent with a patchy or irregular radiomic appearance.

On the other hand, non-progressor lesions demonstrated higher *MCC (Maximal Correlation Coefficient)* and relatively lower values of *Contrast*, *Total Energy*, *Sum Square*, and *Minimum*. Taken together, these features correspond to lesions with a flatter and more homogeneous texture, consistent with a slower rate of change and a less heterogeneous phenotype.

For lymph nodes, a different pattern was observed. High values of *Busyness*—a texture feature that quantifies the rate of intensity changes between adjacent pixels—were strongly predictive of treatment response. Conversely, the combination of low *Busyness* values with elevated *Minimum*, *Energy*, and *Skewness* was associated with the progressor phenotype, indicating a nodal texture characterized by reduced pixel-level intensity variation but accompanied by overall higher signal values and distributional asymmetry.

3.5. Research quality

Radiomics has been a rapidly growing discipline over the past 10 years. In fact, since 2015, there have been 16,315 results in PubMed with the keyword “radiomics” (updated on October 1, 2025).



However, its use in real-world clinical practice is still very limited and requires further clinical validation. The main issue lies in the lack of standardization, which could be facilitated by the introduction of structured guidelines and tools for assessing the quality of studies. Often, the quality of studies involving radiomic features is poor.

In 2017, Lambin et al. introduced the Radiomics Quality Score (RQS), which includes 16 items to evaluate the entire process in a radiomics study, with a total score ranging from -8 to +36. Recently, new guidelines have been introduced to improve the quality of radiomics studies, particularly the CheckList for EvaluAtion of Radiomics Research (CLEAR) and the METhodological RadiomICs Score (METRICS). Specifically, CLEAR aims to promote transparency in studies, while METRICS is another method for testing the methodological quality of radiomics research. Published in 2024, METRICS includes 30 items that evaluate all the stages and scenarios encountered during radiomics research.[33] The METhodological RadiomICs

Score (METRICS) (https://insightsimaging.springeropen.com/articles/10.1186/s13244-023-01572-w)

of our study was 86.1% which corresponds to an “Excellent” Quality Category (*Table1*).

Category	Item/Condition	Description	Weight	Answer
Study Design	Item#1	Adherence to radiomics and/or machine learning-specific checklists or guidelines	0.0368	Yes
Study Design	Item#2	Eligibility criteria that describe a representative study population	0.0735	Yes
Study Design	Item#3	High-quality reference standard with a clear definition	0.0919	Yes
Imaging Data	Item#4	Multi-center	0.0438	Yes
Imaging Data	Item#5	Clinical translatability of the imaging data source for radiomics analysis	0.0292	Yes
Imaging Data	Item#6	Imaging protocol with acquisition parameters	0.0438	Yes
Imaging Data	Item#7	The interval between imaging used and reference standard	0.0292	Yes
Segmentation	Condition#1	Does the study include segmentation?	-	Yes

Segmentation	Condition#2	Does the study include fully automated segmentation?	-	No
Segmentation	Item#8	Transparent description of segmentation methodology	0.0337	Yes
Segmentation	Item#9	Formal evaluation of fully automated segmentation	0.0225	
Segmentation	Item#10	Test set segmentation masks produced by a single reader or automated tool	0.0112	Yes
Image Processing and Feature Extraction	Condition#3	Does the study include hand-crafted feature extraction?	-	Yes
Image Processing and Feature Extraction	Item#11	Appropriate use of image preprocessing techniques with transparent description	0.0622	Yes
Image Processing and Feature Extraction	Item#12	Use of standardized feature extraction software	0.0311	Yes
Image Processing and Feature Extraction	Item#13	Transparent reporting of feature extraction parameters	0.0415	Yes
Feature Processing	Condition#4	Does the study include tabular data?	-	Yes
Feature Processing	Condition#5	Does the study include end-to-end deep learning?	-	No

Feature Processing	Item#14	Removal of non-robust features	0.02	No
Feature Processing	Item#15	Removal of redundant features	0.02	Yes
Feature Processing	Item#16	Appropriateness of dimensionality compared to data size	0.03	No
Feature Processing	Item#17	Robustness assessment of end-to-end deep learning pipelines	0.02	Yes
Preparation for Modeling	Item#18	Proper data partitioning process	0.0599	Yes
Preparation for Modeling	Item#19	Handling of confounding factors	0.03	No
Metrics and Comparison	Item#20	Use of appropriate performance evaluation metrics for task	0.0352	Yes
Metrics and Comparison	Item#21	Consideration of uncertainty	0.0234	No
Metrics and Comparison	Item#22	Calibration assessment	0.0176	Yes
Metrics and Comparison	Item#23	Use of uni-parametric imaging or proof of its inferiority	0.0117	No
Metrics and Comparison	Item#24	Comparison with a non-radiomic approach or proof of added clinical value	0.0293	Yes
Metrics and Comparison	Item#25	Comparison with simple or classical statistical models	0.0176	No
Testing	Item#26	Internal testing	0.0375	Yes

Testing	Item#27	External testing	0.0749	Yes
Open Science	Item#28	Data availability	0.0075	Yes
Open Science	Item#29	Code availability	0.0075	Yes
Open Science	Item#30	Model availability	0.0075	Yes
Total METRICS score: 86.1% | Quality category: Excellent				

Table 1. The METhodological RadiomICs Score (METRICS) for the study

3.6 Comparison with the gold standard

Among the 165 lesions evaluable for PD-L1 expression, assessment was not feasible in 43 cases due to insufficient tissue availability. Of the remaining 122 lesions, 32 exhibited low PD-L1 expression, whereas 90 showed high expression (tumor proportion score, TPS ≥ 50%).

Based on these data, the predictive performance of PD-L1 TPS was evaluated against clinical response. The analysis yielded 40 true positives and 16 true negatives, but also 50 false positives and 16 false negatives, resulting in an overall accuracy of 45.9%. These findings indicate that, within this cohort, PD-L1 TPS alone had limited discriminatory power to reliably identify responders and non-responders.

When compared with the SHAP-based radiomic model, a notable difference emerges. The radiomics-derived features captured patterns of intralesional heterogeneity and intensity variation that were highly predictive of progression and treatment response. In contrast, PD-L1 expression, despite its established role as a clinical biomarker, failed to provide consistent predictive value, misclassifying more than half of the evaluated cases.

This comparison highlights the potential complementary role of radiomics: while PD-L1 reflects a molecular snapshot of tumor biology, radiomic features

provide a non-invasive, whole-lesion characterization that accounts for spatial heterogeneity, which may better align with clinical outcomes. These results suggest that radiomic profiling, either alone or in combination with molecular biomarkers such as PD-L1, could enhance the accuracy of predictive models and support more informed patient stratification in clinical practice.

4.0 Discussion

Radiomics (and its subtypes, delta radiomics analysis) has emerged as a promising approach to extract quantitative information from medical images, providing a potential link between radiological phenotypes and underlying tumor biology. In the context of immunotherapy, radiomics may offer new opportunities to non-invasively predict treatment response and monitor disease evolution, complementing existing clinical and molecular biomarkers. However, despite its growing potential in the few years, several methodological and technical limitations still affect its reproducibility and clinical applicability.

The primary objective of the study was to assess the real-world applicability of radiomics for predicting treatment response in patients receiving immunotherapy, by developing two predictive models for lung lesions and lymph nodes, which were validated thanks to the support of two external datasets. The results of the study highlight both the potential and the challenges of radiomics analysis in the prediction of response to immunotherapy at the lesion level in patients with NSCLC.

Moreover, unlike most previous studies, the analysis has been conducted on a “lesion-level basis”, that is considering each lesion individually, even more than one in the same patient (up to two lesions where selected), due to the

well-documented phenomenon of intra-patient heterogeneity in immunotherapy response patterns.[23,24]

It is important to underline that these machine learning algorithms were based exclusively on radiomic features, without incorporating any clinical variables.[25] While this decision can be considered a limitation due to the restricted amount of information provided to the model, it also highlights that radiomic features alone can offer independent and valuable predictive clinical information. In recent years, several circulating biomarkers have been proposed as prognostic and predictive factors for response to immunotherapy, however it remains unclear whether the addition of relatively simple clinical variables, such as PD-L1 expression level, the neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR), or other biomarkers obtained from liquid biopsy, could further enhance the predictive capacity of the algorithm; for this reason, further research is needed.[25,26]

The internal validation dataset (training set) was first examined, showing that the chosen models achieved excellent predictive performance, with AUC values of 0.89 for lung lesions and 0.95 for lymph nodes. The best performance for classifying lung lesions obtained an F1-score of 0.77, indicating a balanced performance in terms of both precision and recall, the precision was 0.75, suggesting that 75% of predicted progressive cases were correct, while the recall of 0.81 shows that it successfully identified 81% of all true progressive lesions. These metrics indicate that, within the main cohort, radiomic features were able to reliably differentiate progressor from non-progressor lesions, capturing subtle image-based phenotypes that may escape conventional clinical or radiological assessments.

To test the model's generalizability, performance was assessed on two external independent test sets: TestSet1 (Parma) and TestSet2 (Messina). When the models were externally validated, performance decreased but remained consistent across independent datasets. In external testing, the lung model produced an F1-score of 0.69 on both datasets, coupled with perfect recall, while the lymph node model, tested only on TestSet1, achieved a score of 0.77. The AUC values for lung lesions in TestSet1 and TestSet2 were 0.30 and 0.46, respectively, which are considerably lower than the AUC observed in the main cohort. Similarly, the AUPRC for lung lesions in the test sets was respectively 0.41 and 0.58, indicating that while the model was able to identify some progressive lesions, its overall discrimination ability in external data was limited. These results can be determined by the heterogeneity in the test sets, especially in TestSet1 compared to the main cohort.

TestSet2 consisted of 26 lung lesions of patients enrolled at the Papardo Hospital in Messina, whose segmentation was performed by the same radiologists who segmented the CT scans of the training set. Additionally, the CT scans in TestSet2 included only the portal phase of image acquisition. These similarities made the first validation cohort highly comparable to the main cohort. Consequently, the reproducibility of the test could be mistakenly interpreted as an indication that this cohort was “easier” for the model to characterize.

In contrast the CT scans in TestSet1 were more heterogenous in the imaging protocols and segmentation. TestSet1, with 61 lesions, came from the cohort of patients treated at the University Hospital of Parma. In this case, the lesions were selected and segmented by an entirely different team of radiologists, with different years of experience and obtained on a different TC scan with

different protocol acquisitions (some lesions were evaluated in arterial phase and a few with non-contrast phase), however the segmentation criteria and software were equivalent. Despite the decrease in performance from internal to external testing, the models demonstrated comparable results across both TestSets, demonstrating generalizability across institutions, imaging protocols, segmentation workflows, and radiologist experience.

Notably, one recurring trend observed was the tendency of the classifiers of the model to over-predict progression (successfully identifying over 80% of progressive cases in both datasets), as reflected in the relatively high recall but lower precision values. This bias has a dual interpretation. From a methodological standpoint, it highlights a weakness, as it increases the number of false positives. From a clinical perspective, however, it may represent an advantage: prioritizing sensitivity ensures that most progressive lesions are identified, which is crucial in a setting where missing progression could have significant therapeutic implications. Physicians could therefore use such a model to flag lesions at high risk of progression, directing early local interventions such as stereotactic radiotherapy while maintaining systemic immunotherapy.

The most notable result is the consistency of predictive performance across both external validation cohorts, supporting the robustness of the models. An interesting observation was that the predictive model developed for lymph nodes appeared more accurate and reproducible, despite having a smaller sample size, which typically poses limitations in machine learning. This could be attributed to the inner characteristics of lymph node lesions, such as sharper margins and regular shape and absence of adjacent anatomical (air-density structures, such as major bronchi), and pathological structures (e.g.,

atelectatic lung parenchyma), which likely make segmentation easier and more reliable.

Although radiomics has proven capable of detecting molecular alterations in cancer, it cannot capture disease at the cellular level or microenvironment because it analyzes voxels of nearly millimetric dimensions. Instead, it assesses the phenotypes that certain alterations present on radiological images. The SHAP-based feature analysis further reinforced the biological plausibility of the models, unveiling two distinct radiomic phenotypes: one phenotype correlated with unresponsiveness to immunotherapy and another phenotype associated with a positive response to immunotherapy. Progressor lung lesions were characterized by high *Total Energy*, *Energy*, and *Contrast*, features indicative of intralesional heterogeneity and patchy areas of high attenuation (indicating vivid contrast enhancement). This radiomic profile may correspond to immune-excluded tumors, where stromal components or poorly inflamed tumor regions limit immunotherapy efficacy. In contrast, the second phenotype of non-progressor lesions exhibited a more homogeneous and "flat" appearance, with higher *MCC* and lower *Energy* and *Contrast*. This pattern resembles an immuno-inflamed phenotype, more likely to respond to immune checkpoint inhibition. In lymph nodes, a similar differentiation emerged: lesions with high *Busyness* (reflecting rapid changes in local pixel intensity) were associated with treatment response, while progressors displayed low *Busyness* combined with higher *Minimum*, *Energy*, and *Skewness*. Together, these findings suggest that radiomics can capture image-based surrogates of tumor-immune microenvironment interactions.

The algorithm demonstrated limited predictive accuracy; in fact, it is unable to correctly predict the treatment response in approximately 3 out of 10 patients**.** However, it is important to note that these results are comparable

to those of PD-L1 tissue expression analysis evaluation on tissue, with the key difference being that our approach does not require a tissue sample from every lesion. In fact, the key strength of the study lies in the comparison with the gold standard biomarker, PD-L1 expression.[31] In our cohort, PD-L1 TPS achieved an accuracy of only 45.9%, misclassifying more than half of the cases. This result is consistent with clinical evidence showing that PD-L1, while useful, is far from a perfect biomarker due to intra-tumoral heterogeneity, temporal variability, and sampling limitations.[32] Radiomics, by contrast, provided a non-invasive, whole-lesion characterization, it is less expensive and exploitable to any lesion in the same patient, offering higher predictive accuracy and the ability to account for spatial heterogeneity across multiple lesions in the same patient. This comparison underlines the complementary role of radiomics: while molecular assays provide mechanistic insights at the cellular level, radiomics integrates structural and textural heterogeneity at the macroscopic level, potentially bridging the gap between tumor biology and clinical outcomes.

Although this study boasts a high METRICS score (see table above) and the encouraging findings, some limitations must be acknowledged. Primarily, the cohort size was relatively small, which reduces statistical power and may limit the model's capacity to learn from diverse imaging phenotypes. Additionally, we employed semi-automatic/manual segmentation conducted by different operators. Although reflecting real-world clinical practice, contributing to the model's generalizability, it could also have introduced unnecessary variability, for instance inter-operator variability. Future integration of fully automated segmentation could further improve reproducibility and reduce bias, enhancing the results of the study.

Another point of interest is the lesion-level approach itself. By intentionally evaluating up to two lesions per patient, it was possible to capture intra-patient heterogeneity of immunotherapy response, a phenomenon increasingly recognized in clinical oncology. While this strategy reflected the real-world complexity of tumor behavior and strengthens the clinical relevance of the models, it also introduces a degree of statistical dependence, as lesions from the same patient may share biological characteristics, tumor microenvironmental conditions, or imaging features.

Clinically, this approach supports the concept of lesion-level decision-making, wherein local treatments (e.g., stereotactic radiotherapy) may be selectively applied to non-responding lesions while systemic therapy is continued.

A key intrinsic limitation of radiomics is the need for an accurate segmentation, and as such, some lesions may not be segmentable due to intrinsic imaging segmentation challenges (e.g., imaging artifacts, atelectasis, closeness to the pulmonary hila etc.).

A larger training set would provide greater heterogeneity, which in turn would improve the model's ability to recognize a wider range of cases. While some of the limitations mentioned, such as the application of test-retest methods, are hindered by ethical or technical challenges, others can be more easily addressed. The inclusion of clinical variables—such as circulating biomarkers, gender, ethnicity, smoking status, BMI, and others—could enhance the algorithm's performance. For instance, previous works demonstrated that the use of extracellular vesicles (EVs) PD-L1, might increase the accuracy of radiomics in advanced NSCLC treated with single agent immunotherapy. Likely the use of circulating tumor DNA (ctDNA) dynamics, which demonstrated promising activity as predictive biomarker during PD-1/PD-L1 blockage and is a more reliable biomarker than EVs might be implemented in future studies.

Similarly, increasing the number of patients in the training set would also contribute to improving the model.

In the framework of these findings and study limitations, new avenues are emerging to further enhance the integration of radiomics into real-world clinical practice. The successful clinical implementation of radiomics requires not only robust algorithms but also continuously expanding and well-defined databases that reflect the heterogeneity of real patient populations, as already explained. This entails the systematic creation of patient subpopulations, defined by both clinical and molecular characteristics, and the adoption of increasingly precise and standardized segmentation workflows. These elements are fundamental to improving reproducibility, a key factor for the clinical reliability of radiomic models.

To address these needs, since July 2024, we have undertaken a structured expansion of our radiomic database with the objective of improving the reproducibility of both segmentation and feature extraction. Specifically, a retrospective database including 90 patients diagnosed with small-cell lung cancer (SCLC) has been established, in which segmentation of the primary lesion (often very challenging for the central location and local invasion, typical of this type of lung cancer) and any pathological lymph nodes was performed.

Furthermore, since beginning of 2025 the project has been further extended to include all lesions that underwent CT-guided biopsy at the Department of General and Interventional Radiology (Dr. Giuseppe Cittadini) in IRCCS San Martino Polyclinic Hospital, over the past ten years. This expanded dataset comprises approximately 400 biopsy-proven lesions, encompassing a wide spectrum of pathologies such as primary lung lesions, secondary (metastatic) lesions, and benign entities. Pending the integration of fully automated

segmentation tools, lesion delineation is currently being performed manually on the CT scans acquired on the day of the biopsy by radiologists with different levels of experience. This approach not only enables the assessment of inter-operator reproducibility but also provides a valuable and heterogeneous imaging resource for future radiomic analyses and potential multi-institutional collaborations.

In conclusion, this study demonstrates that radiomic analysis can predict lesion-level response to immunotherapy with promising accuracy and reproducibility, outperforming PD-L1 in our dataset. Although limited by cohort size and manual segmentation, the models showed consistent performance across independent cohorts, supporting their generalizability. Although this study has limitations, many of these can be addressed in near future research. The decision not to predict overall patient survival was deliberate, as such an endpoint is influenced by multiple confounding factors. Lung cancer patients often present with frailty and comorbidities, and the analysis of a single lesion cannot, at least for now, provide conclusive information regarding survival outcomes, which are affected by variables such as extra-thoracic metastatic lesions, performance status scale (ECOG scale), and age.[35] For these reasons, the study focused on lesion-level treatment response, a more reliable and clinically meaningful endpoint that may directly inform patient management. Predicting the progression of an individual lesion could, in the future, guide decisions such as whether to locally irradiate that lesion to achieve better disease control. In earlier disease settings, such as neoadjuvant immunotherapy, radiomics-based response prediction could assist multidisciplinary Disease Management Teams in determining the optimal treatment approach. Radiomic analysis may soon be integrated into

clinical practice, enhancing the personalization of treatment for NSCLC patients.

In conclusion, our study supports the rapid translation of radiomics into clinical practice. While tailored treatments have established the field of personalized medicine, radiomics, with its ability to provide a comprehensive view of the tumor, has the potential to extend these boundaries and lead to a "super-personalized" medicine, where treatment is adapted not only to the patient but to the biological behavior of each individual lesion.[36]

References

1. Smolarz B, Łukasiewicz H, Samulak D, Piekarska E, Kołaciński R, Romanowicz H. Lung Cancer—Epidemiology, Pathogenesis, Treatment and Molecular Aspect (Review of Literature). International Journal of Molecular Sciences. 2025; 26(5):2049. https://doi.org/10.3390/ijms26052049
2. Ilie M, Juco J, Huang L, Hofman V, Khambata-Ford S, Hofman P. Use of the 22C3 anti-programmed death-ligand 1 antibody to determine programmed death-ligand 1 expression in cytology samples obtained from non-small cell lung cancer patients. Cancer Cytopathol. 2018;126(4):264-274. doi:10.1002/cncy.21977
3. Hwang DM, Albaqer T, Santiago RC, et al. Prevalence and Heterogeneity of PD-L1 Expression by 22C3 Assay in Routine Population-Based and Reflexive Clinical Testing in Lung Cancer. J Thorac Oncol. 2021;16(9):1490-1500. doi:10.1016/j.jtho.2021.03.028
4. Frank C. Detterbeck, Gavitt A. Woodard, Anna S. Bader, Sanja Dacic, Michael J. Grant, Henry S. Park, Lynn T. Tanoue, The Proposed Ninth Edition TNM Classification of Lung Cancer, CHEST, Volume 166, Issue 4, 2024
5. Zhang C, Zhu S, Yuan Y, Dai S. Comparison Between Endobronchial-Guided Transbronchial Biopsy and Computed Tomography-Guided Transthoracic Lung Biopsy for the Diagnosis of Central Pulmonary Lesions. Clin Respir J. 2024 Sep;18(9):e70015. doi: 10.1111/crj.70015. PMID: 39314190; PMCID: PMC11420531.
6. Lu S, Stein JE, Rimm DL, et al. Comparison of Biomarker Modalities for Predicting Response to PD-1/PD-L1 Checkpoint Blockade: A Systematic Review and Meta-analysis. JAMA Oncol. 2019;5(8):1195-1204. doi:10.1001/jamaoncol.2019.1549
7. Ulas EB, Hashemi SMS, Houda I, Kaynak A, Veltman JD, Fransen MF, Radonic T, Bahce I. Predictive Value of Combined Positive Score and Tumor Proportion Score for Immunotherapy Response in Advanced NSCLC. JTO Clin Res Rep. 2023 May 25;4(9):100532. doi: 10.1016/j.jtocrr.2023.100532. PMID: 37681219; PMCID: PMC10480627.

8. Genova C, Tasso R, Rosa A, et al. Prognostic Role of Soluble and Extracellular Vesicle-Associated PD-L1, B7-H3 and B7-H4 in Non-Small Cell Lung Cancer Patients Treated with Immune Checkpoint Inhibitors. Cells. 2023;12(6):832. doi:10.3390/cells12060832
9. Coroller TP, Agrawal V, Huynh E, et al. Radiomic-Based Pathological Response Prediction from Primary Tumors and Lymph Nodes in NSCLC. J Thorac Oncol. 2017;12(3):467-476. doi:10.1016/j.jtho.2016.11.2226
10. Wu S, Zhan W, Liu L, et al. Pretreatment radiomic biomarker for immunotherapy responder prediction in stage IB-IV NSCLC (LCDigital-IO Study): a multicenter retrospective study. J Immunother Cancer. 2023;11(10):e007369. doi:10.1136/jitc-2023-007369
11. Chen M, Lu H, Copley SJ, et al. A Novel Radiogenomics Biomarker for Predicting Treatment Response and Pneumotoxicity From Programmed Cell Death Protein or Ligand-1 Inhibition Immunotherapy in NSCLC. J Thorac Oncol. 2023;18(6):718-730. doi:10.1016/j.jtho.2023.01.089
12. Trebeschi S, Drago SG, Birkbak NJ, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. Ann Oncol. 2019;30(6):998-1004. doi:10.1093/annonc/mdz108
13. Fave, X. et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. Sci. Rep. 7, 588 (2017).
14. Khorrami M, Prasanna P, Gupta A, et al. Changes in CT Radiomic Features Associated with Lymphocyte Distribution Predict Overall Survival and Response to Immunotherapy in Non-Small Cell Lung Cancer. Cancer Immunol Res. 2020;8(1):108-119. doi:10.1158/2326-6066.CIR-19-0476
15. Prelaj A, Miskovic V, Zanitti M, et al. Artificial intelligence for predictive biomarker discovery in immuno-oncology: a systematic review. Annals of Oncology. 2024;35(1):29-65. doi:10.1016/j.annonc.2023.10.125
16. Yan K, Zhang D. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sensors and Actuators B: Chemical. 2015;212:353-363. doi:10.1016/j.snb.2015.02.025

17. Suthaharan S. Support Vector Machine. In: Suthaharan S, ed. Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning. Springer US; 2016:207-235. doi:10.1007/978-1-4899-7641-3_9
18. K-nearest neighbor. ResearchGate. Accessed July 30, 2025. https://www.researchgate.net/publication/220580323_K-nearest_neighbor
19. Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32. doi:10.1023/A:1010933404324
20. Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. RUSBoost: Improving classification performance when training data is skewed. In: 2008 19th International Conference on Pattern Recognition. ; 2008:1-4. doi:10.1109/ICPR.2008.4761297
21. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair. 2002;16:321-357. doi:10.1613/jair.953
22. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17. Curran Associates Inc.; 2017:4768-4777.
23. Saito Y, Horiuchi S, Morooka H, et al. Inter-tumor heterogeneity of PD-L1 expression in non-small cell lung cancer. J Thorac Dis. 2019;11(12):4982-4991. doi:10.21037/jtd.2019.12.24
24. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. Eur Radiol. 2012;22(4):796-802. doi:10.1007/s00330-011-2319-8
25. Rossi G, Barabino E, Fedeli A, et al. Radiomic Detection of EGFR Mutations in NSCLC. Cancer Res. 2021;81(3):724-731. doi:10.1158/0008-5472.CAN-20-0999
26. Moding EJ., Shahrokh Esfahani M., Jin C, Hui AB, Nabet BY, Liu Y, Chabon JJ, Binkley MS, Kurtz DM, et al.. Integrating ctDNA Analysis and Radiomics for Dynamic Risk Assessment in Localized Lung Cancer. Cancer Discov. 2025 Aug

4;15(8):1609-1629. doi: 10.1158/2159-8290.CD-24-1704. PMID: 40299851; PMCID: PMC12324966.

27. Cella, E.; Zullo, L.; Marconi, S.; Rossi, G.; Coco, S.; Dellepiane, C.; Alama, A.; Rozeboom, L.; Bennicelli, E.; Parisi, F.; et al. Immunotherapy-Chemotherapy Combinations for Non-Small Cell Lung Cancer: Current Trends and Future Perspectives. Expert Opin Biol Ther 2022, 22, 1259–1273, doi:10.1080/14712598.2022.2116273.
28. Forde,.M.; Spicer, J.; Lu, S.; Provencio, M.; Mitsudomi, T.; Awad, M.M.; Felip, E.; Broderick, S.R.; Brahmer, J.R.; Swanson, S.J.; et al. Neoadjuvant Nivolumab plus Chemotherapy in Resectable Lung Cancer. N Engl J Med 2022, 386, 1973–1985, doi:10.1056/NEJMoa2202170.
29. Felip, E.; Altorki, N.; Zhou, C.; Csőszi, T.; Vynnychenko, I.; Goloborodko, O.; Luft, A.; Akopov, A.; Martinez-Marti, A.; Kenmotsu, H.; et al. Adjuvant Atezolizumab after Adjuvant Chemotherapy in Resected Stage IB-IIIA Non-Small-Cell Lung Cancer (IMpower010): A Randomised, Multicentre, Open-Label, Phase 3 Trial. Lancet 2021, 398, 1344–1357, doi:10.1016/S0140-6736(21)02098-5.
30. Genova, C.; Boccardo, S.; Mora, M.; Rijavec, E.; Biello, F.; Rossi, G.; Tagliamento, M.; Dal Bello, M.G.; Coco, S.; Alama, A.; et al. Correlation between B7-H4 and Survival of Non-Small-Cell Lung Cancer Patients Treated with Nivolumab. J Clin Med 2019, 8, doi:10.3390/jcm8101566.
31. Rossi, G.; Barabino, E.; Fedeli, A.; Ficarra, G.; Coco, S.; Russo, A.; Adamo, V.; Buemi, F.; Zullo, L.; Dono, M.; et al. Radiomic Detection of EGFR Mutations in NSCLC. Cancer Res 2021, 81, 724–731, doi:10.1158/0008-5472.CAN-20-0999.
32. 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 (Basel) 2022, 14, 350, doi:10.3390/cancers14020350.
33. Kocak B, Akinci D'Antonoli T, Mercaldo N, et al. METhodological RadiomICs Score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMII. Insights Imaging. 2024 doi:10.1186/s13244-023-01572-w

34. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017 Dec;14(12):749-762. doi: 10.1038/nrclinonc.2017.141. Epub 2017 Oct 4. PMID: 28975929.
35. Budin-Ljosne, I. et al. DataSHIELD: an ethically robust solution to multiple-site individual-level data analysis. Public Health Genomics 18, 87–96 (2015).
36. MAASTRO clinic. euroCAT: Distributed Learning for Individualized Medicine. youtube.com. http://youtu.be/ ZDJFOxpwqEA. (2014).