PET/CT radiomics and machine learning enable non-invasive survival stratification and histologic tumor risk profiling in patients with lung adenocarcinoma

Meixin Zhao

Peking University Third Hospital

Kilian Kluge

Medical University of Vienna Department of Biomedical Imaging and Image-guided Therapy:
Medizinische Universität Wien Universitätsklinik für Radiologie und Nuklearmedizin

Laszlo Papp

Medical University of Vienna: Medizinische Universität Wien

Marko Grahovac

Medical University of Vienna: Medizinische Universität Wien

Shaomin Yang

Peking University Health Science Centre

Chunting Jiang

Peking University Third Hospital Department of Nuclear Medicine

Denis Krajnc

Medical University of Vienna: Medizinische Universität Wien

Clemens P. Spielvogel

Medical University of Vienna: Medizinische Universität Wien

Boglarka Ecsedi

Medical University of Vienna: Medizinische Universität Wien

Alexander Haug

Medical University of Vienna Department of Biomedical Imaging and Image-guided Therapy:
Medizinische Universität Wien Universitätsklinik für Radiologie und Nuklearmedizin

Shiwei Wang

Evomics Medical Technology Co.,Ltd

Marcus Hacker

Medical University of Vienna Department of Biomedical Imaging and Image-guided Therapy:
Medizinische Universität Wien Universitätsklinik für Radiologie und Nuklearmedizin

Weifang Zhang

Peking University Third Hospital Department of Nuclear Medicine

Xiang Li (✉ lixiangmed@gmail.com)

Research Article

Keywords: Lung adenocarcinoma, PET/CT, Radiomics, Machine learning, Survival prediction, Histology prediction

Posted Date: August 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-771161/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose

Risk stratification in patients with lung adenocarcinoma (LUAD) is mandatory for treatment guiding and outcome prediction. Amongst clinical parameters including histological analyses, imaging procedures provide important information. The present study aimed to investigate the ability of machine learning models trained on clinical and 2-deoxy-2-[^{18}F]fluoro-D-glucose (^{18}F FDG) positron emission tomography/computed tomography (PET/CT) derived radiomic data to predict overall survival (OS), tumor grade (TG), and histologic growth pattern risk (GPR) in treatment naïve LUAD patients.

Methods

421 treatment naïve patients with histologically diagnosed lung adenocarcinoma and available ^{18}F FDG PET/CT imaging were retrospectively analyzed. Four patient cohorts were generated based on the available data for predicting 4-year OS ($n = 276$), 3-year OS ($n = 280$), TG ($n = 298$) and GPR ($n = 265$). ^{18}F FDG-positive lesions were delineated semiautomatically, from which 2082 radiomic features were extracted and combined with endpoint-specific clinical and demographic parameters. Machine learning models were built for the prediction of 4-year OS ($M_{4\text{OS}}$), 3-year OS ($M_{3\text{OS}}$), tumor grading (M_{TG}) and histologic growth pattern risk (M_{GPR}), respectively. Monte Carlo (MC) cross-validation with 100-folds and 80:20 training to validation split was employed as a performance evaluation for all models. Kaplan-Meier survival analysis was performed to assess the association between the $M_{4\text{OS}}$ and $M_{3\text{OS}}$ predictions with OS.

Results

Area under the receiver operator characteristics curve (AUC) was highest for $M_{4\text{OS}}$ (AUC 0.88), followed by $M_{3\text{OS}}$ (AUC 0.84), while M_{TG} and M_{GPR} performed equally well (AUC 0.76). Predictions of $M_{4\text{OS}}$ (HR 0.128, $p < 0.000001$) and $M_{3\text{OS}}$ (HR 0.0942, $p < 0.000001$) were independently associated with OS.

Conclusion

In our retrospective cohorts, machine learning models demonstrated the ability to prognosticate long-term survival outcomes in patients with lung adenocarcinoma. Furthermore, tumor lesions could be characterized according to their histologic grade and predominant growth pattern risk with high accuracy.

Declarations

Funding: This study was funded by the National Major Science and Technology Projects of China (CN) (2016YFC0103705) and the Key Clinical Project of Peking University Third Hospital “Construction of lymphoma prediction model based on PET/CT radiomics and machine learning” (BYSYZD2019038).

Conflicts of interest: The authors have no conflict of interest to declare.

Availability of data and material: The collected data files are available on request from the corresponding author.

Authors’ contributions:

Author	Contribution
Meixin Zhao	All data collection, imaging data and patient data analysis
Kilian Kluge	Manuscript writing, Literature research, Review
Laszlo Papp	Radiomics and machine learning methodological design and execution, review
Marko Grahovac	Data set preparation, Data preprocessing
Shaomin Yang	Evaluation of pathological results
Chunting Jiang	Patients outcome evaluation
Denis Krajnc	Data set preparation, Data preprocessing
Clemens P. Spielvogel	MUW Radiomic Engine validation, Survival analysis
Boglarka Ecsedi	Implementation and validation of the MUW Radiomics Engine according to IBSI guidelines and reference datasets
Alexander Haug	Supervision, review
Shiwei Wang	Technical support for machine learning analysis
Marcus Hacker	Supervision, review
Weifang Zhang	Co-supervision of the patient management
Xiang Li	Supervision of the study, design of the concept, review

Ethics approval: This study was approved by the ethics committee of Peking University Third Hospital (LM2020001).

Acknowledgments: Laszlo Papp is the founder of the Dedicaid-Evomics Automated Machine Learning platform (Dedicaid GmbH, Vienna, & Evomics Medical Technology Co., Ltd., Shanghai).

Introduction

Non-small cell lung cancer (NSCLC) accounts for 85% of all primary lung cancers. Despite recent and continued advances in screening, diagnosis and treatment, NSCLC continues to be the leading cause of cancer-related mortality worldwide [1] with a dreary 5-year relative survival rate of 4% for patients with advanced disease [2]. NSCLCs most common subtype is lung adenocarcinoma (LUAD), making LUAD the most prevalent histologic type of all primary lung cancers in both sexes [3]. The prognosis of patients with LUAD and equal TNM stages differed significantly depending on their progressive features and predominant growth patterns [4–7], which was considered in the 2015 published world health organization (WHO) classification of lung tumors [8]. However, histopathologic workup for diagnosis is challenging in inoperable patients, and it is frequently and solely based on small tumor samples from biopsy or cytology [9]. Due to the histologically heterogeneous nature of LUADs, these are prone to sampling errors and clinical complications [10,11]. Hence a complementary, ideally non-invasive method for improved prognostication and tumor characterization is highly desirable.

Hybrid imaging with [^{18}F]FDG PET/CT as a non-invasive method is well established in the initial staging [12,13] and proved useful in response assessment and follow-up [14]. Novel post-processing methods for PET/CT imaging data enable high-throughput quantitative feature extraction from biomedical images known as radiomics and offer a non-invasive approach to capture intra-tumoral heterogeneity [15], while supervised machine learning provides the tools to map high-dimensional input data to investigational endpoints effectively [16]. The potential of PET and CT-derived radiomic features in combination with machine learning has been demonstrated inter alia in lesion characterization [17–19], response prediction to various therapeutic options [20,21] and survival prognostication [20,22]. However, few studies have studied LUAD cohorts and those who did often exhibited methodological shortcomings like lack of appropriate data preprocessing.

We hypothesized that [^{18}F]FDG PET/CT-derived radiomic features in combination with clinical data would be predictive of survival outcomes and histologic tumor risk profiles in LUAD patients.

The study objective was to assess the ability of machine learning models trained on clinical and [^{18}F]FDG PET/CT derived radiomic data to predict 4- and 3-year overall survival (OS), tumor grade (TG), and histologic growth pattern risk (GPR) in treatment naïve LUAD patients.

Materials And Methods

Patient selection and study design

A total of 421 treatment-naïve patients (215 female, mean 63 ± 11 years SD, range 28-91) who got whole-body [^{18}F]FDG-PET/CT between 01/04/2012 and 30/04/2019 were included in this single-center

retrospective study based on the following criteria. All patients had histologically confirmed adenocarcinoma of the lung.

Patients with severe image artifacts in PET/CT or concomitant pneumonia, atelectasis, pleural adhesions, and other cancer types were excluded. For all patients, clinical data such as age, sex, TNM stage, predominant growth pattern, degree of tumor differentiation, EGFR mutation status, smoking status and treatment strategy were gleaned from their medical records if available or inferred from primary clinical data such as medical imaging data, surgical or histological records.

Tumor stages were defined according to the American Joint Committee on Cancer Staging Manual [23]. The definition of growth patterns and tumor grading was based on the classification of lung adenocarcinomas published by the IASLC/ATS/ERS in 2011 [24]. Growth pattern risk stratification was done by defining micropapillary (MMP), solid predominant (SPA) and variants of invasive (VIA) lung adenocarcinomas as high-risk and all other growth patterns as low-risk lesions following the literature on the prognostic subgroups of lung adenocarcinomas [4–7]. All inconclusive histologic samples were subsumed in others. Histopathologic tumor grades were dichotomized into poor and moderate to high differentiation. Patients were split into four cohorts (4-year OS, 3-year OS, TG, GPR) based on the available data on the predicted endpoints (see Table 1).

The primary endpoints of this study were overall survival, which was defined as the time from PET/CT examination until death or the date of the last contact with the patient. The minimum follow-up period to ascertain 4-year and 3-year OS were 0.07; the maximum was 87.85 months (30.1 ± 20.4 , 29.8 ± 21.1 median and SD months respectively). The secondary endpoints were the predictions of tumor grading and histologic growth pattern risk.

Ethical approval was obtained and the need for informed consent was waived by the medical ethics committee of the Peking University Third Hospital, Beijing, People's Republic of China.

FDG PET/CT image acquisition

All [^{18}F]FDG-PET/CT studies were obtained on a Siemens Biograph TruePoint PET/CT system. Prior to tracer injection, all patients had been fasting for a minimum of 6 h. Measured blood glucose values in all patients were 104 ± 16 mg/dl. Patients were intravenously injected with 5.55 MBq/kg (0.15 mCi/kg) [^{18}F]FDG and rested recumbent in a calm environment afterwards. Sixty minutes post-injection, the scans were obtained from skull base to upper femur. PET scans were acquired in 5-7 bed positions with 2 minutes per bed position followed by deep inspiratory high-resolution CT scans at 120 kVp and 100 mAs without intravenous contrast. PET images were reconstructed iteratively using the ordered-subset expectation maximization (OSEM) algorithm (21 subsets, 3 iterations) and subsequently CT-data-based scatter and attenuation corrected (PET matrix size 168x168, CT matrix size 512x512).

Texture analysis

Initial delineation of the volume of interests (VOI) of the primary and secondary tumor lesions was performed semiautomatically using the Hybrid 3D software ver. 4.0.0 (Hermes Medical Solutions, Stockholm, Sweden) with delineation criteria set according to the PET Response Criteria in Solid Tumors (PERCIST 1.0) recommendations for target lesions at baseline [25], mediastinal blood pool as background and minimal lesion size greater or equal to 130 voxels. Lesions smaller than 130 voxels or with subthreshold uptake were defined manually based on the CT images. Subsequently, delineations were reviewed and if needed, adjusted manually according to a consensus decision of two nuclear medicine experts. Finally, kriging interpolation was used to resample all delineated VOIs to 2.0 x 2.0 x 2.0 uniform voxel resolution [26].

Following the Imaging Biomarker Standardization Initiatives (IBSI) guidelines, radiomic features with “very strong” and “strong” consensus values were extracted from each individual lesion using the Medical University Wien (MUW) Radiomic Engine (ver. 2.0), which was prior validated according to IBSI standards [27] (see Supplemental Material - Table 1a). To obtain lesion number-independent feature vectors of equal length per patient, lesions of the same patient were ordered by their volume and the distribution of each radiomic feature across the ordered lesions was determined, resulting in a distribution function for each radiomic feature. The distribution function then underwent IBSI intensity histogram feature evaluation, generating overall 2082 radiomic features. Lastly, radiomic features, number of lesions, patient characteristics and varying clinical characteristics depending on the predicted endpoint were merged to form feature vectors. (see Figure 2)

Cross-validation and machine learning scheme

Monte Carlo (MC) cross-validation, data preprocessing and ensemble machine learning was performed using the Dedicaid Automated Machine Learning platform (Dedicaid GmbH, Vienna). First a MC cross-validation scheme was applied assigning the patients in each cohort randomly into training and validation sets 100 times with a ratio of 80-20%, resulting in 100 unique folds per cohort. Equal subsampling of validation samples ensured balanced representation within the validation set. (see Supplemental Material; Sections 4YOS, 3YOS, GPR, TG). Subsequently all folds underwent preprocessing by feature range normalization, feature redundancy reduction, feature selection and sample balancing. Feature redundancy reduction was performed by covariance matrix analysis defining redundancy as an absolute Pearson correlation coefficient greater than 0.85 [28]. Feature selection followed the rules of the curse of dimensionality, where number of features were selected (S = training sample count) in each fold. To avoid class imbalance Synthetic Minority Over-sampling Technique (SMOTE) was employed to synthesize new instances of the minority class between preexisting class members [29].

Machine learning (ML) was performed by a mixed, stacked ensemble learning scheme to build predictive models of 4-year overall survival, 3-year overall survival, tumor grade and growth pattern risk (models abbreviated as M_{4OS} , M_{3OS} , M_{TG} , M_{GPR} respectively). (see Supplemental Table 3a-d and 4a-d). Kaplan-Meier survival analysis was performed based on the M_{4OS} and M_{3OS} predictions using the Lifelines Python package.

Results

Patients

In the tumor grade cohort, 237/298 (80%) specimens were of moderate or high differentiation, while 61 (20%) were poorly differentiated. In the growth pattern risk cohort, the most common histologic subtypes were APA (n = 126, 47%), followed by SPA (n = 53, 20%) and LPA (n = 39, 15 %). Regarding inherent histologic subtype risks, 200/265 (76%) tumors were categorized as low risk and 65 (24%) as high risk (see Table 1).

Feature extraction, ranking and extraction

MC cross-validation scheme assignment of the 276, 280, 298, 265 patients of the 4-year OS, 3-year OS, tumor grade and growth pattern risk cohorts, respectively, resulted in 100 unique folds with 221, 224, 238, 212 samples in each of the respective training and 55, 56, 60, 53 in each of the respective validation sets. Subsequent preprocessing by SMOTE and covariance matrix analysis-based redundancy reduction resulted in 328, 270, 450, 374 respective samples and 36, 32, 42, 38 respective features across all folds. (Supplemental Material – Section 4YOS, 3YOS, TG, GPR).

Model performances and validation

The cross-validated M_{4OS} , M_{3OS} , M_{TG} , M_{GPR} model performances across all MC-folds was established by confusion matrix analysis and yielded in average 80%, 78%, 82%, 59% sensitivity, 88%, 85%, 55%, 84% specificity, 88%, 84%, 65%, 80% positive predictive value, 82%, 80%, 77%, 68% negative predictive value and 84%, 82%, 69%, 71% accuracy, respectively (see Figure 3). The mean Area under the Receiver Operator Characteristics curve (AUC) of all built ensemble models was 0.88, 0.84, 0.76, 0.76, respectively (see Figure 2 and Supplemental Material – Table 7a-d).

The survival predictions of M_{4OS} (HR 0.128, $p < 0.000001$) and M_{3OS} (HR 0.0942, $p < 0.000001$) were independently associated with OS (see Figure 4)

Feature weighting and distribution

Out of the 30 most predictive features of the 4-year overall survival prediction, 19 were PET radiomic features and 11 CT radiomic features. No clinical features were selected. The three most predictive feature families were NGLDM (n = 9), GLRLM (n = 7) and GLCM (n = 5). In the 3-year overall survival prediction, PET (n = 14) and CT radiomic features (n = 16) were nearly equally distributed. The three most predictive feature families were GLCM (n = 9), followed by Histogram and Morphological features (n = 6 each). Out of the 30 most predictive features of the tumor grade prediction, 25 CT radiomic features and 5 PET radiomic features were selected. The three most predictive feature families were GLCM (n = 9), followed by GLSZM and Intensity features (n = 6 each). In the growth pattern risk prediction, 29 features were PET radiomic features plus the degree of differentiation feature. The most predictive feature

families were intensity derived ($n = 9$), followed by GLCM, GLRM and NGLDM features ($n = 4$ each). (see Figure 5 and Supplemental Material – Table 8a-d)

Discussion

In this study, we developed predictive models of long-term survival outcomes and histologic characteristics of treatment naïve patients with LUAD based on [^{18}F]FDG PET/CT-derived radiomics features and clinical data. We demonstrated excellent cross-validated model performances for the survival endpoints ($M_{40\text{S}}$ and $M_{30\text{S}}$) and satisfying cross-validated model performances for the histologic endpoints (M_{TG} and M_{GPR}).

The prognosis of LUAD patients varies depending on their predominant histologic growth pattern [4,5]. Often histologic information is only obtainable from needle biopsy specimens [9] which are prone to high rates of sampling bias [10,11]. Yet, accurate prognostication is key to optimal follow-up planning and disease management in oncologic diseases.

Radiomic approaches offer the possibility of non-invasive lesion characterization and have shown promising results in survival prognostication in NSCLC cohorts [20,30,31]. Literature on the role of radiomics in survival predictions in pure lung adenocarcinoma cohorts is relatively scarce [32,33]. Yang et al. [32] studied the performances of a PET/CT-radiomics based- (Concordance-Index (C-Index) 0.737), a clinical-data based- (C-Index 0.729) and a composite model (C-Index 0.803) in predicting overall survival in patients with EGFR positive LUADs. They discovered that CT derived simple first- and second-order statistical GLCM- and GLSZM-based textural features were most predictive. Similarly, Choe et al. [33] compared CT-based radiomics models with and without clinical features (C-Indices were 0.805 and 0.764, respectively) to a clinical model (C-Index 0.735) in predicting overall survival in LUAD patients after resection. Our PET/CT radiomics based overall survival prediction models $M_{40\text{S}}$ (AUC 0.88; HR 0.128, $p < 0.000001$) and $M_{30\text{S}}$ (AUC 0.84; HR 0.0942, $p < 0.000001$) demonstrated excellent cross-validated performances and hence higher predictive performances compared to the previous studies [32,33]. In line with Yang et. al. [32] first-order histogram-based and GLCM-based statistical textural features were the ones predominantly informative for our $M_{40\text{S}}$ and $M_{30\text{S}}$ models. Conversely to their findings, PET-derived features were more or equally prognostic compared to CT-derived ones for our survival models. Interestingly, clinical features did not contribute to either of our survival models. The superiority in predictive performances and discrepancy concerning the principal informative modality might be caused by our novel approach to radiomics feature extraction from multiple lesions by the construction of a feature distribution function across lesions resulting in novel derivative features. Our demonstrated high-performance in $M_{30\text{S}}$ and $M_{40\text{S}}$ models implies that this type of analysis should be considered as an alternative multi-lesion radiomics analysis approach next to the popular merging method, which handles all disjunctive VOI as one virtual VOI in the 3D space [34].

The secondary endpoints of our study were histological predictions of tumor grading and growth pattern risk groups. Both cross-validated model performances of M_{TG} and M_{GPR} demonstrated satisfactory

performances with AUCs of 0.76 each. To the best of our knowledge, no previous study has predicted lung adenocarcinoma tumor grading by means of PET/CT radiomics. Bae et. al [35] studied the CT radiomics and conventional PET parameters inter alia as a potential biomarker for tumor grades in 91 LUADs and reported excellent predictive performances for grade one through three. However, their study was based on a relatively small and highly imbalanced cohort (Grade 1 – n = 19, Grade 2 – n = 65, Grade 3 – n = 7) without employing data balancing techniques, leading to biased results. They found mostly simple histogram-based and textural uniformity features significantly correlated with tumor grades, which is in line with our findings. PET/CT radiomic based histologic growth pattern predictions were studied by Shao et. al. [18] in 91 preoperative stages I LUAD patients. They reported an AUC of 0.804 for the prediction of lesion membership to the intermediate-high growth pattern risk group. Again, their study was based on a small and highly imbalanced cohort (low-risk n = 18, intermediate-high risk n = 75) without employing data balancing techniques, leading to biased results. In line with their findings [18], first-order histogram-based and second-order statistical texture features were predominantly informative for our M_{GPR} model. Contrary to their results, PET-derived features were the only predictive features in our model. Surprisingly, M_{GPR} was the only model in which a non-radiomic feature was ranked as informative to the prediction. This finding stands in contrast to the general trend of the incremental predictive power of composite models over pure radiomic models [30,32,33].

Our study was limited by its retrospective and single-center design lacking external validation.

However, we mitigated some of those limitations by employing a MC-fold scheme keeping training and validation samples strictly apart to avoid information leakage. Furthermore, data preprocessing (feature normalization, redundancy reduction, feature selection and class balancing) as well as mixed ensemble machine learning and 100-fold cross-validation were used to reduce false discoveries.

Conclusion

In our retrospective cohort, machine learning models demonstrated the ability to prognosticate long-term survival outcomes in patients with lung adenocarcinoma. Furthermore, tumor lesions could be characterized according to their histologic grade and predominant growth pattern risk with high accuracy.

References

1. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature*. 2018;553:446–54.
2. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. *CA: A Cancer Journal for Clinicians*. 2019;69:363–85.
3. Curado MP, Edwards B, Shin HR. *Cancer Incidence in Five Continents*. IARC Scientific Publications; 2007.

4. Yoshizawa A, Motoi N, Riely GJ, Sima CS, Gerald WL, Kris MG, et al. Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: Prognostic subgroups and implications for further revision of staging based on analysis of 514 stage i cases. *Modern Pathology* Nature Publishing Group. 2011;24:653–64.
5. Warth A, Muley T, Meister M, Stenzinger A, Thomas M, Schirmacher P, et al. The novel histologic International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification system of lung adenocarcinoma is a stage-independent predictor of survival. *J Clin Oncol*. 2012;30:1438–46.
6. Strand TE, Rostad H, Strøm EH, Hasleton P. The percentage of lepidic growth is an independent prognostic factor in invasive adenocarcinoma of the lung. *Diagnostic Pathology* Diagnostic Pathology. 2015;10:1–7.
7. Li H, Cao W. Pulmonary enteric adenocarcinoma: A literature review. *Journal of Thoracic Disease*. 2020;12:3217–26.
8. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances since the 2004 Classification. *Journal of Thoracic Oncology*. 10: International Association for the Study of Lung Cancer; 2015. pp. 1243–60.
9. Loo PS, Thomas SC, Nicolson MC, Fyfe MN, Kerr KM. Subtyping of undifferentiated non-small cell carcinomas in bronchial biopsy specimens. *Journal of Thoracic Oncology* International Association for the Study of Lung Cancer. 2010;5:442–7.
10. Soler Cataluña JJ, Perpiñá M, Greses J, Calvo V, Padilla JD, París F. Cell type accuracy of bronchial biopsy specimens in primary lung cancer. *Chest*. 1996;109:1199–203.
11. Huang K-Y, Ko P-Z, Yao C-W, Hsu C-N, Fang H-Y, Tu C-Y, et al. Inaccuracy of lung adenocarcinoma subtyping using preoperative biopsy specimens. *The Journal of Thoracic and Cardiovascular Surgery*. 2017;154.
12. Ettinger DS, Wood DE, Aggarwal C, Aisner DL, Akerley W, Bauman JR, et al. Non-small cell lung cancer, version 1.2020: Featured updates to the NCCN guidelines. *JNCCN Journal of the National Comprehensive Cancer Network* [Internet] Harborside Press. 2019;17:1464–72.
13. Planchard D, Popat S, Kerr K, Novello S, Smit EF, Faivre-Finn C, et al. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2018;29:iv192–237.
14. Cuaron J, Dunphy M, Rimner A. Role of FDG-PET scans in staging, response assessment, and follow-up care for non-small cell lung cancer. *Front Oncol*. 2013;2 JAN:1–8.
15. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–6.
16. Papp L, Spielvogel CP, Rausch I, Hacker M, Beyer T. Personalizing medicine through hybrid imaging and medical big data analysis. *Front Phys*. 2018;6:1–19.

17. Sha X, Gong G, Qiu Q, Duan J, Li D, Yin Y. Identifying pathological subtypes of non-small-cell lung cancer by using the radiomic features of 18F-fluorodeoxyglucose positron emission computed tomography. *Translational Cancer Research*. 2019;8:1741–9.
18. Shao X, Niu R, Shao X, Jiang Z, Wang Y. Value of 18F-FDG PET/CT-based radiomics model to distinguish the growth patterns of early invasive lung adenocarcinoma manifesting as ground-glass opacity nodules. *EJNMMI Res*. 2020 Jul 13;10(1):80.
19. Chang C, Sun X, Wang G, Yu H, Zhao W, Ge Y, et al. A Machine Learning Model Based on PET/CT Radiomics and Clinical Characteristics Predicts ALK Rearrangement Status in Lung Adenocarcinoma. *Front Oncol*. 2021;11:1–12.
20. Kirienko M, Cozzi L, Antunovic L, Lozza L, Fogliata A, Voulaz E, et al. Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery. *European Journal of Nuclear Medicine Molecular Imaging European Journal of Nuclear Medicine Molecular Imaging*. 2018;45:207–17.
21. Valentinuzzi D, Vrankar M, Boc N, Ahac V, Zupancic Z, Unk M, et al. [18F]FDG PET immunotherapy radiomics signature (iRADIOMICS) predicts response of non-small-cell lung cancer patients treated with pembrolizumab. *Radiol Oncol*. 2020;54:285–94.
22. Kirienko M, Sollini M, Corbetta M, Voulaz E, Gozzi N, Interlenghi M, et al. Radiomics and gene expression profile to characterise the disease and predict outcome in patients with lung cancer. *European Journal of Nuclear Medicine and Molecular Imaging*; 2021.
23. Edge SB, Compton CC. The american joint committee on cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010;17:1471–4.
24. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *Journal of Thoracic Oncology*. 2011;6:244–85.
25. Joo Hyun O, Lodge MA, Wahl RL. Practical percist: A simplified guide to PET response criteria in solid tumors 1.0. *Radiology [Internet]*. 280: Radiological Society of North America Inc.; 2016. pp. 576–84.
26. Stytz MR, Parrott RW. Using kriging for 3d medical imaging. *Computerized Medical Imaging and Graphics [Internet]*. Comput Med Imaging Graph. 1993;17:421–42.
27. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. 2016 [cited 2021 Apr 24];.
28. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology [Internet]*. Radiological Society of North America Inc.; 2016 [cited 2021 Apr 25];278:563–77. Available from: <https://pubmed.ncbi.nlm.nih.gov/26579733/>.
29. Amin A, Anwar S, Adnan A, Nawaz M, Howard N, Qadir J, et al. Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study.
30. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non—Small Cell Lung Cancer. *Radiology*.

2016;281:947–57.

31. Mu W, Jiang L, Zhang JY, Shi Y, Gray JE, Tunali I, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. *Nat Commun.* 2020;16(1):5228. 11(.
32. Yang B, Ji H, Zhong J, Ma L, Zhong J, Dong H, et al. Value of 18F-FDG PET/CT-Based Radiomics Nomogram to Predict Survival Outcomes and Guide Personalized Targeted Therapy in Lung Adenocarcinoma With EGFR Mutations. *Frontiers in Oncology.* Frontiers Media S.A.; 2020. ;10.
33. Choe J, Lee SM, Do KH, Kim S, Choi S, Lee JG, et al. Outcome prediction in resectable lung adenocarcinoma patients: value of CT radiomics. *European Radiology* *European Radiology.* 2020;30:4952–63.
34. Haider SP, Mahajan A, Zeevi T, Baumeister P, Reichel C, Sharaf K, et al. PET/CT radiomics signature of human papilloma virus association in oropharyngeal squamous cell carcinoma. *European Journal of Nuclear Medicine and Molecular Imaging.* 2020;47.
35. Bae JM, Jeong JY, Lee HY, Sohn I, Kim HS, Son JY, et al. Pathologic stratification of operable lung adenocarcinoma using radiomics features extracted from dual energy CT images. *Oncotarget.* 2017;8:523–35.

Tables

Table 1: Demographic and clinical patient characteristics

Patient characteristic	4-year overall survival		3-year overall survival		Tumor grade		Growth pattern risk	
	n	%	n	%	n	%	n	%
Total patients	276		280		298		265	
Mean age ± SD	70 ± 11		65 ± 11		62 ± 11		62 ± 11	
Sex								
female	140	51	140	50	163	55	147	56
male	136	49	140	50	135	45	118	44
TNM stage								
0	3	1	3	1	6	2	6	2
1	84	30	84	30	141	47	139	53
2	7	3	7	2	14	5	14	5
3	47	17	47	17	48	16	45	17
4	135	49	139	50	89	30	61	23
Predominant growth pattern								
AIS	2	1	3	1	6	2	6	2
MIA	5	2	5	2	10	3	9	3
LPA	19	7	19	7	39	13	39	15
APA	77	28	77	27	126	42	126	47
PPA	11	4	11	4	20	7	20	8
MPP	3	1	2	1	5	2	5	2
SPA	37	13	38	14	53	18	53	20
VIA	3	1	4	1	7	2	7	3
unknown	119	43	121	43	32	11		
Growth pattern risk								
low risk	115	41	115	41	201	67	200	76

high risk	43	16	11	16	65	22	65	24
unknown	118	43	121	43	32	11		
Tumor grade								
poor differentiation	59	22	61	22	61	20	35	13
moderate-high differentiation	139	50	140	50	237	80	230	87
unknown	78	28	79	28	-	-	-	-
Tissue invasion								
no tissue invasion	3	1	3	1	7	2	7	3
microinvasion	5	2	6	2	10	3	9	3
tissue invasion	268	97	271	97	281	95	249	94
EGFR mutation status								
positive	108	39	108	39	171	57	157	59
negative	95	34	95	34	127	43	108	41
unknown	78	28	77	27				
Source of specimen								
resection	95	34	95	34	175	59	173	65
needle biopsy	181	66	185	66	123	41	92	35
Smoking status								
Never smoker	127	46	130	46	179	60	165	62
History of smoking	83	30	84	30	72	24	61	23
unknown	66	24	66	24	47	16	39	15
Treatment received after imaging								
Surgery	97	35	97	35	178	60	176	66
Radiotherapy	24	9	24	9	29	9,7	35	13
Chemotherapy	124	45	125	45	110	37	92	35

Targeted therapy	104	38	106	38	108	36	95	36
------------------	-----	----	-----	----	-----	----	----	----

Data are *n*; % percentage; age in years; SD – standard deviation

Figures

Figure 1

Conceptual data extraction and analysis framework. A [¹⁸F]FDG-PET/CT scan lesion delineation B Extraction of radiomics feature from all lesions and combination with patients clinical characteristics to a feature vector C Assignment of patients feature vectors to training and validation sets by a Monte Carlo cross-validation scheme with 20%:80% validation to training split ratio and 100 unique folds. D Subsequent data preprocessing of all folds by means of normalization, data set balancing, feature sorting as well as selection and redundancy reduction. E Training of ensemble machine learning scheme models (L1 Layer 1, L2 Layer 2) and cross-validation of their results. F Analysis of predictive performances across all folds by confusion matrices and Area under the Receiver operator curves as well as Kaplan-Meier survival analysis to assess association between the M40S and M30S predictions with OS

Figure 2

Study flow diagram. Through an institutional database search 421 patients were retrospectively recruited. Survival-, histopathologic-, staging-, tumor extent data (see Table 1) was gleaned from medical records. Patients were split into four cohorts (4-, 3-year OS, TG and GPR) based on the available data on the predicted endpoints. Primary and secondary thoracic lesions were delineated on PET/CTs. Radiomic features with “strong” and “very strong” IBSI consensus values were extracted from all lesions. A radiomic

feature distribution function for each patient was constructed which underwent subsequent feature analysis resulting in 2082 radiomic features per patient. The number of lesions per patient, radiomic features, patient characteristics (sex, age) and endpoint specific clinical characteristics were used to construct the final feature vectors. Monte Carlo (MC) cross-validation scheme was utilized to assign patients to training and validation roles. Subsequent data preprocessing ensured normalized features, feature redundancy reduction and balanced classes in all folds. Predictive models were built by ensemble machine learning scheme. Predictive performance across all folds was analyzed by confusion matrices and Area under the Receiver operator curves. Kaplan-Meier survival analysis was performed to assess association between the M4OS and M3OS predictions

Figure 3

Mean cross-validation Receiver Operator Characteristics (ROC) curves of a) M4OS b) M3OS c) MDD d) MGPR. The thick blue line corresponds to the mean ROC curve, while the light blue shaded area represents the spread of all 100 ROC curves generated across the validation folds. AUC – Area Under the Receiver Operator Characteristics Curve (mean of all 100 AUCs); FPR – False Positive Rate, TPR – True Positive Rate. Dashed diagonal line represents a reference random-guess AUC for comparison

Figure 4

Box plots showing the cross-validated performance of the top-layer (MLL3) prediction of a) 4-year OS, b) 3-year OS, c) Tumor grade and d) Growth pattern risk. Performance values were established by confusion matrix analysis across all MC folds. SNS – Sensitivity; SPC – Specificity; PPV – Positive Predictive Value; NPV – Negative Predictive Value; ACC – Accuracy; Performance values as percentages (%)

Figure 5

Kaplan-Meier curves according to the predictions of the a) M4OS b) M3OS. A highly significant association between the model predictions and overall survival was shown for both models.

Figure 6

Bar plots showing the feature family membership of the 30 most informative features as number of features (n) per feature family according to their derived imaging modality and clinical origin. a) 4-year OS, b) 3-year OS, c) Tumor grade and d) Growth pattern risk. CT – computed tomography; PET – positron emission tomography; GLCM – grey level co-occurrence matrix; NGLDM - neighboring grey level dependence matrix; NGTDM - neighborhood grey tone difference matrix, GLSZM - grey level size zone matrix; GLDZM - grey level distance zone matrix; GLRLM - gray level run length matrix, IV - intensity histogram, DD – Degree of Differentiation

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [LungCancerSupplement.docx](#)