Original Study

Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers

Giovanni Visonà,[1] Lisa M. Spiller,[2] Sophia Hahn,[2] Elke Hattingen,[2,3,4,5] Thomas J. Vogl,[3,4,5,6] Gabriele Schweikert,[7] Katrin Bankov,[8] Melanie Demes,[3,4,5,8] Henning Reis,[3,4,5,8] Peter Wild,[3,4,5,8] Pia S. Zeiner,[3,4,5,9] Fabian Acker,[3,4,5,10] Martin Sebastian,[3,4,5,10] Katharina J. Wenger[2,3,4,5]

Abstract

We trained and validated machine learning models to identify Non–small-cell lung cancer patients with a high risk of developing brain metastases, as they could potentially benefit from surveillance brain magnetic resonance imaging. Early detection of asymptomatic brain metastases is crucial to improve clinical prospects. Employed prospectively at initial diagnosis, such models can help select high-risk subgroups.

Purpose: Non–small-cell lung cancer (NSCLC) shows a high incidence of brain metastases (BM). Early detection is crucial to improve clinical prospects. We trained and validated classifier models to identify patients with a high risk of developing BM, as they could potentially benefit from surveillance brain MRI. **Methods:** Consecutive patients with an initial diagnosis of NSCLC from January 2011 to April 2019 and an in-house chest-CT scan (staging) were retrospectively recruited at a German lung cancer center. Brain imaging was performed at initial diagnosis and in case of neurological symptoms (follow-up). Subjects lost to follow-up or still alive without BM at the data cut-off point (12/2020) were excluded. Covariates included clinical and/or 3D-radiomics-features of the primary tumor from staging chest-CT. Four machine learning models for prediction (80/20 training) were compared. Gini Importance and SHAP were used as measures of importance; sensitivity, specificity, area under the precision-recall curve, and Matthew's Correlation Coefficient as evaluation metrics. **Results:** Three hundred and ninety-five patients compromised the clinical cohort. Predictive models based on clinical features offered the best performance (tuned to maximize recall: sensitivity∼70%, specificity∼60%). Radiomics features failed to provide sufficient information, likely due to the heterogeneity of imaging data. Adenocarcinoma histology, lymph node invasion, and histological tumor grade were positively correlated with the prediction of BM, age, and squamous cell carcinoma histology were negatively correlated. A subgroup discovery analysis identified 2 candidate patient subpopulations appearing to present a higher risk of BM (female patients + adenocarcinoma histology, adenocarcinoma patients + no other distant metastases). **Conclusion:** Analysis of the importance of input

[1]Empirical Inference, Max-Planck Institute for Intelligent Systems, Tübingen, Germany
[2]Goethe University Frankfurt, University Hospital, Institute of Neuroradiology, Frankfurt am Main, Germany
[3]University Cancer Center Frankfurt (UCT), Frankfurt am Main, Germany
[4]Frankfurt Cancer Institute (FCI), Frankfurt am Main, Germany
[5]German Cancer Research Center (DKFZ) Heidelberg, Germany and German Cancer Consortium (DKTK), Partner Site Frankfurt, Mainz, Germany
[6]Goethe University Frankfurt, University Hospital, Department of Diagnostic and Interventional Radiology, Frankfurt am Main, Germany
[7]Division of Computational Biology, School of Life Sciences, University of Dundee, Dundee, UK
[8]Goethe University Frankfurt, University Hospital, Dr. Senckenberg Institute of Pathology, Frankfurt am Main, Germany
[9]Goethe University Frankfurt, University Hospital, Edinger Institute, Institute of Neurology, Frankfurt am Main, Germany
[10]Goethe University Frankfurt, University Hospital, Department of Medicine II, Hematology/Oncology, Frankfurt am Main, Germany

Submitted: Jan 11, 2023; Revised: Jul 24, 2023; Accepted: Aug 1, 2023; Epub: xxx

Address for correspondence: Katharina J. Wenger, MD, MHBA, Goethe University Frankfurt, University Hospital, Institute of Neuroradiology, Schleusenweg 2-16, Frankfurt am Main 60528, Germany.
E-mail contact: katharina.wenger@kgu.de

https://doi.org/10.1016/j.cllc.2023.08.002

Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002

features suggests that the models are learning the relevant relationships between clinical features/development of BM. A higher number of samples is to be prioritized to improve performance. Employed prospectively at initial diagnosis, such models can help select high-risk subgroups for surveillance brain MRI.

Clinical Lung Cancer, Vol. 000, No.xxx, 1–12 © 2023 The Author(s). Published by Elsevier Inc.

Keywords: NSCLC, Interpretable machine learning, Predictive models, Radiomics, Secondary brain cancer

Introduction

Lung cancer is one of the most commonly diagnosed cancers worldwide with an estimated 2.2 million cases in 2020 alone and is the leading cause of cancer death.[1] With approximately 85% of all lung cancers, Non–small-cell lung cancer (NSCLC) summarizes a heterogeneous group of histologies. The World Health Organization (WHO) has classified 3 main types: squamous cell carcinoma, adenocarcinoma, and large-cell carcinoma.[2] With 20% to 40% of NSCLC patients, the incidence of brain metastases (BM) is one of the highest in cancer subtypes.[3-7] Certain genetic aberrations seem to be related to an increased risk. It is estimated that approximately 30% to 70% of patients with epidermal growth factor receptor (EGFR) mutations,[6,8,9] 60% to 70% of patients with anaplastic lymphoma kinase (ALK) rearrangements[10-12] and 30% to 40% of patients with proto-oncogene tyrosine-protein kinase 1 (ROS1) rearrangements[12,13] develop BM of the course of their disease. The incidence rates vary with the type of study, patient selection methods, and follow-up intervals.

BM are associated with a high rate of morbidity and mortality which make their early detection and personalized treatment important clinical goals.[7,14] The European Society for Medical Oncology (ESMO) guidelines currently state that if available, brain imaging with magnetic resonance imaging (MRI) or computed tomography (CT) should be carried out at initial diagnosis, but is most relevant in patients with neurological symptoms or signs.[15,16] The American College of Radiology (ACR) recommends MRI of the brain at initial diagnosis in any patient with clinical stage II, III, or IV even in the absence of neurologic symptoms and at any stage in case of neurological symptoms.[17] However, routine surveillance brain MRI for patients without neurological symptoms is not advised.[15,18] Brain scans of patients at initial clinical staging identify BM in 3% to 10% of patients.[19-21] It is well known, that there is a considerable number of patients with asymptomatic BM at initial presentation and during follow-up.[22-26] While prophylactic cranial irradiation (PCI) treatment can reduce the incidence of BM by approximately 50%, current data suggests that it fails to improve overall survival (OS) and may lead to a cognitive decline.[27-29] Early detection of BM is therefore crucial for the timely initiation of local therapy and systemic treatment with central nervous system (CNS) penetrance. Local therapeutic options for BM currently include surgical interventions, stereotactic radiotherapy, and whole brain radiation therapy (WBRT), depending on the size and number of lesions, location, and on the patient's performance status before or in conjunction with.[30,31]

In this work, we trained and validated classifier models to identify NSCLC patients with a high risk of developing BM over the course of their disease, as they could potentially benefit from surveillance brain MRI. The models were trained on a selected set of clinical features such as age, sex, or initial stage of cancer, as well as tumor histology and genetic features. Additionally, we examined the use of 3D-radiomics-features of the primary tumor from staging chest-CT. The semi-automatic segmentation pipeline was deployed using an End-to-end artificial intelligence (AI) solution for medical research. The prediction models trained offer insights on the challenges involved in the classification task described. The primary aim is the development of a sufficiently robust model to support prospective studies by highlighting patients who are more likely to develop BM. To the best of our knowledge, no such study has been conducted on an unselected NSCLC patient collective.[32]

Material and Methods

Patients

This retrospective study was approved by the Institutional Review Board (IRB) of the University Hospital blinded (UCT-3-2018). Consecutive patients with initial diagnosis of NSCLC from January 2011 to April 2019 and in-house chest-CT scan (staging) were identified in the blinded clinical cancer registry. In this centralized data collection, all cancer patients from our hospital and referring hospitals are fully documented as inpatients and outpatients, with documentation of their history, diagnostic workup, and follow-up information. The tumor documentation was done by trained medical documentation specialists and documentation assistants. Patients were followed-up until December 20th 2020 (data cut-off). Patients underwent brain imaging at initial diagnosis (contrast-enhanced MRI; only if there were absolute contraindications, contrast-enhanced CT) and in case of neurological symptoms or signs during follow-up.

Clinical Feature Extraction and Data Processing

The clinical parameters extracted from the clinical cancer registry with a data cut-off on December 20th 2020 are listed in Table 1. Indication for testing for driver mutations changed over the recruitment period. If patient's informed consent was given and biomaterials (tumor tissue) were available in the Interdisciplinary Biobank and Database blinded (iBDF), testing was completed for this study.

We excluded patients without BM who were lost to follow-up or still alive without BM at the data cut-off date (48/443). Patients were classified as "alive without BM" if they were still in routine surveillance at our hospital (intervals/frequency and modality of surveillance imaging dependent on stage/prior treatment) without neurological symptoms or if so with a negative MRI scan in line with guideline recommendations.

Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002

Table 1 The Clinical Parameters Extracted From the Clinical Cancer Registry if Available

Patient characteristics
Age at initial diagnosis
Sex
Date of death (DOD)
Tumor characteristics
UICC TNM categories at initial diagnosis
UICC TNM stage at initial diagnosis
Histological grading at initial diagnosis
BM at initial diagnosis (yes/no)
BM over the course of disease (yes/no)
Other distant metastases at initial diagnosis (yes/no)
ICD-O Code at initial diagnosis
EGFR mutation (mutated with Exon/wild-type)
BRAF mutation (mutated/wild-type)
ALK rearrangement (EML4-ALK fusion protein; positive/negative)
ROS 1 rearrangement (positive/negative)
PD-L1 expression (positive/negative)[a]

In approximately 20% of cases the primary tumor was not bioptable and therefore, tissue analysis was conducted on distant metastases.
[a] With regard to PD-L1 expression, tumor proportion score (TPS) was documented: Percentage of viable tumor cells showing partial or complete membrane staining at any intensity, relative to all viable tumor cells; 0 = 0%-1%; 1 = 1-<5%; 2 = 5-<10%; 3 = 10-<25%; 4 = 25-<50%, 5 =>50%.[74] A score of 1 or higher was considered positive.

In a first step for the processing of the raw data, we removed the uninformative features, as well as features that were missing for an excessive number of patients. This included the information on 4 out of the 5 mutations/gene expression signatures analyzed (EML4_ALK, PD_L1, ROS-1, BRAF), which were missing for more than 75% of the patients. Details on the missing values imputation for all features are given in Supplemental File 3 and Supplemental Table S1. We removed features that directly contained information on the presence of BM to avoid information leakage, such as the UICC TNM stage or the UICC TNM M category. In order to preserve non-BM-related information from the M category, a binary indicator of the presence of other distant metastases was introduced. To avoid excessive stratification, we mapped each ICD-O code to a category consisting of its first 3 digits, which are indicative of the histology of the tumor. The information on BM consists of 2 features, indicative of the presence of BM at first diagnosis and the development of BM during the course of the disease. As the classification target, we used a binary feature with a value of 1 if the patient had developed BM at any stage of the disease, and 0 otherwise. This process produced a dataset with 8 clinical features and a corresponding set of binary labels. For the classification analysis, the ICD-O 3-digits feature was 1-hot encoded, which results in the substitution of that feature with 11 binary features. The final set of clinical features consisted of 18 features. We first performed a preliminary analysis to confirm previously established correlations of clinical features with the development of BM.[4,33-35] To this end, we considered these 18 features and employed Kendall's Tau Hypothesis Testing for a direct comparison with the binary variable indicating the presence of BM. We did not employ any form of data imputation for this analysis. To correct for multiple hypothesis testing, we employed the Benjamini-Hochberg method[36] implemented in the Python package multiply.[37]

Technical Parameters for CT Image Acquisition

All patients underwent chest contrast-enhanced CT, mostly in addition to nonenhanced CT as first-line imaging investigation of NSCLC. Imaging was performed at 2 different institutions (Institute of Diagnostic and Interventional Radiology and Institute of Neuroradiology, both University Hospital Frankfurt am Main, Germany) on 6 different scanners, as hardware was subject to changes over time. The region of interest (ROI) covered the gross tumor volume on mediastinal and lung window. Technical parameters are summarized in Supplemental Table S2.

Radiomics Feature Extraction and Data Processing

The computer-assisted detection (CAD) pipeline for automated segmentation consisted of 3 prominent building blocks: (1) multi-threshold candidate detection, (2) feature extraction, and (3) learning-based nodule classification. A hierarchical support vector machine (SVM) was applied for nodule detection. Details on the cascaded SVM classifier have been published previously.[38] The pipeline was deployed using the Philips IntelliSpace Discovery 3.0, an End-to-end AI solution for medical research (Philips Healthcare, Hamburg, Germany). Correct ROI identification and segmentation were reviewed by an experienced clinical radiologist (K.J.W. > 8 years of experience in oncologic imaging). In case of automated segmentation errors, Simplicit was used for manual correction. Simplicit is a set of tools for 3D interactive segmentation developed in Philips Research Medisys by Jean-Michel Rouet. A detailed description can be found in Supplemental File 1. Pyradiomics v3.0.1, an open-source python package for the extraction of radiomics features from medical imaging, was used for radiomic feature extraction.[39] Patients were excluded from the radiomic analysis in case we were unable to distinguish tumor tissue from nontumor tissue on bolus-enhanced CT even with simultaneous consideration of all additional information available (38/395). Features were extracted from nonenhanced lung windows if available (229/357). Volume segmentation spanning multiple slices was converted into one volume (NIFTII) before extracting features. 3D shape descriptors were extracted and single values per feature for a region of interest were calculated (segment-based). The parameter file for customized extraction (yaml structured) is provided in Supplemental File 2. These features were filtered to remove non-numerical features and features with 0 variances, resulting in a set of 1231 features. In 128 patients, lung windows were reconstructed from IV contrast-enhanced raw data only, which resulted in a significant covariate shift. Several methods have been proposed to account for distributional shifts (such as ComBat[40]), however, they rely on strict assumptions on the type of distribution shift. None of the methodologies tested could ensure that a trained classifier would truly predict the target variable, rather than learning the difference between the 2 IV contrast sets. We therefore filtered the radiomics data to include only patients for whom no IV contrast agent was used (229 samples). Additionally, hierarchical clustering

Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002

Table 2 Description of the Selected Clinical Features

Feature	Number Missing	Description
Distant metastases	0 (0)	Binary indicator of the presence of other distant metastases (than BM) at initial diagnosis
Age	0 (0)	Age at first diagnosis
Sex	0 (0)	Binary indicator for the sex of the patient
Grade	140 (151)	Histological grading at initial diagnosis
TNM_T	3 (3)	UICC TNM category T at initial diagnosis that describes the primary tumor site and size
TNM_N	6 (7)	UICC TNM category N at initial diagnosis that describes the regional lymph node involvement
EGFR	170 (199)	Binary indicator of the presence of mutations in the EGFR gene
ICD-O 3-digits	0 (0)	First 3 digits of the ICD-O code

The parentheses following the number of missing values for each feature represent the number of missing values in the complete set of patients before the removal of the 48 patients with unknown status for the presence of brain metastases.

of the correlation matrix of the radiomics features reveals a high degree of collinearity (Supplemental Figure S1).

Machine-Learning-Based Classification of BM

To verify the feasibility of the prediction task, we trained a series of binary classifiers using as input the set of clinical features (Table 2), radiomics features, or a combination of both. As the radiomics features presented a high degree of collinearity (as evidenced in Supplemental Figure S1), we pre-processed them with a standardization followed by PCA, selecting the number of components to capture 90% of the variance in the data (Supplemental Figure S2). The models selected include Logistic Regression (LR), Random Forests (RF),[41] LightGBM,[38] and the AutoML framework Autogluon.[42] Gini Importance (GI) and the TreeExplainer implementation of SHapley Additive exPlanations (SHAP) were used as measures of importance for the RF model. Deliberation on the choice of models and feature importance is summarized in Supplemental File 3. To measure the classification performance, we chose sensitivity, specificity, the area under the precision-recall curve, and Matthew's Correlation Coefficient (MCC) as evaluation metrics. Given the limited size of the dataset, we evaluated our models using a methodology similar to the nested cross-validation or "double cross" approach.[43] Specifically, we performed 30 random partitions of the data into 80/20 training to test split. If required by the model, the training set is used to fit a K Nearest Neighbor (KNN; using 5 nearest neighbors), which is then employed to fill in missing values in the full dataset. With each training set, we performed a random hyperparameter search using 5-fold cross-validation. The best resulting configuration is selected to train a model on the full training split, and this model is subsequently used to predict labels for the test set. In addition, a similar setup was tested with the AutoML framework Autogluon Tabular.[42] Autogluon Tabular was employed for the clinical data, the radiomics features, and a combination of both.

Identification of Patient Subpopulations

Identifying groups of NSCLC patients that are particularly susceptible to developing BM is crucial to improve clinical care and correctly allocate the necessary resources. More generally, this data mining task is referred to as subgroup discovery, for which many methodologies have been proposed.[44,45] We employed the beam search algorithm implemented in the Python package pysubgroup[46] to identify feature combinations of interest for their link to BM.

Results

Patient Characteristics

The study population selection is depicted in Figure 3. Of 443 retrospectively enrolled patients, 48 patients were lost to follow-up or still alive without BM at the data cut-off point and therefore excluded. The remaining 395 patients compromised our clinical cohort (Table 3). Of these, 43 (11%) were stage I, 19 (5%) stage II, 60 (15%) stage III, and the majority with 273 (69%) stage IV. Most patients were male (63%). The minimum age was 24 years, the maximum 89 and the mean 65. The histologic distribution included 245 patients (62%) with adenocarcinoma, 110 (28%) with squamous cell neoplasia, 18 (5%) with epithelial neoplasia, and 22 (6%) with other histologies. Most patients had no node involvement (104) or regional node involvement (N2, 153). A total of 40% of female patients and 27% of male patients presented with BM at initial diagnosis (32% overall) and 45% of female patients and 33% of male patients with BM at any time during the course of disease. The patient collective was predominantly Caucasian.

Due to the heterogeneity of the existing imaging data, the radiomics cohort compromised only 229 patients. The reasons for the heterogeneity are discussed in detail in the limitations section.

Age and Histology are Correlated With BM

Selecting a false discovery rate (FDR) threshold of 0.01, 3 out of 18 features showed a statistically significant dependence with the presence of BM. The results are summarized in Supplemental Table S3. ICD-O histology codes starting with 814 indicate adenocarcinoma histology and were positively correlated with the presence of BM at any stage of the disease.[33] ICD-O histology codes starting with 807 indicate squamous cell carcinoma histology and showed a reduced incidence of BM compared to nonsquamous tumors.[4] Finally, age displayed a negative correlation with BM.[34] In Figures 1 and 2, we visualize the differences in counts or distributions in these 3 features for patients stratified by the development of BM.

Machine-Learning-Based Classification of BM

Classification models that employed only clinical features as input showed the best performance when tuned to maximize recall to

Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002

Table 3 Patient Characteristics of the Clinical Cohort (n = 395)

Characteristics		n (Unless Otherwise Specified)
Age		65 y (mean; SD 10; range 24-89)
Gender	Male	249 (63%)
	Female	146 (37%)
Histological subtype	Adenocarcinoma	245 (62%) Female 104 (71% of female patients) Male 141 (57% of male patients)
	Squamous cell neoplasia	110 (28%) Female 29 (20% of female patients) Male 81 (33% of male patients)
	Epithelial neoplasia	18 (5%)
	Other	22 (6%)
Histological grading at first diagnosis	1	8
	2	167
	3	79
	4	1
	Unknown	140
EGFR mutation	Mutated	Total 26 Female 14 (10% of female patients tested for mutation) Male 12 (5% of male patients tested for mutation) Exon 19 with highest frequency
	Wildtype	199
	Not performed	170
UICC TNM category N at initial diagnosis that describes the regional lymph node involvement	N0	104
	N1	48
	N2	153
	N3	84
	Nx	6
UICC TNM stage at initial diagnosis	I	43 (11%)
	II	19 (5%)
	III	60 (15%)
	IV	273 (69%)
BM at initial diagnosis		Total 127 (32%) Female: 59 (40% of female patients) Male: 68 (27% of male patients) Adenocarcinoma: 99 (78% of patients with BM at first diagnosis)
BM at any time during the course of disease		148 (38%) Female: 66 (45% of female patients) Male: 82 (33% of male patients) Adenocarcinoma: 114 (77% of patients with BM at any time)
Other distant metastases at first diagnosis (other than BM)		208 (47%)

Of 443 retrospectively enrolled patients, 48 patients were lost to follow-up or still alive without BM at the data cut-off point and therefore excluded.

improve prospective studies, obtaining at best a sensitivity of ~ 70% and a specificity just above 60%. Radiomics features performed suboptimally. The results of the ML-based classification task are presented in Figure 4 and Table 4. They highlight how the use of clinical data appears to be much more informative than the radiomics features available. This observation, however, is skewed by the different number of samples available to the 2 sets (395 samples in clinical dataset vs. 229 samples in radiomics dataset due to heterogeneity of existing imaging data). Indeed, neither classifiers trained on the clinical data for the same subset of patient nor

Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002

Figure 1 Relationship between patient age and development of BM. There is a difference of 6.8 years between the peaks of the distributions of patients with and without BM.

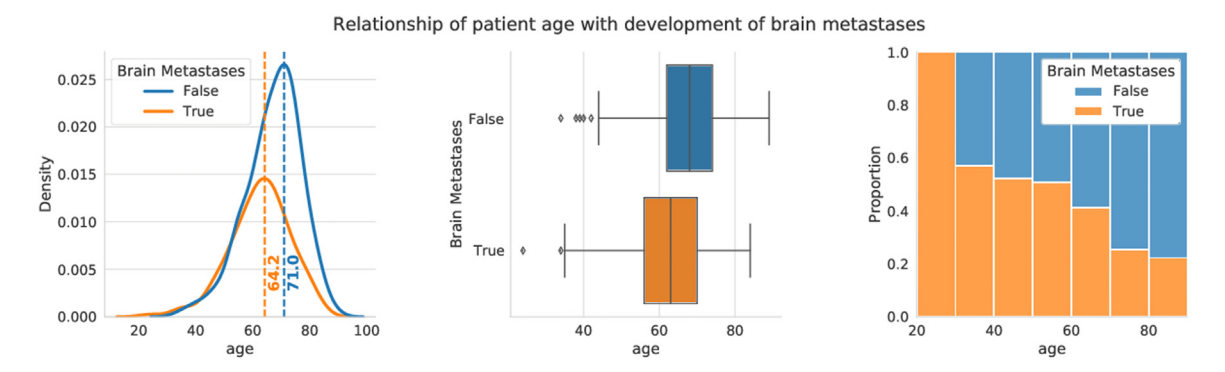


Figure 2 Relationship of the 2 discrete features that rejected the null hypothesis of the Kendall's Tau test with the development of brain metastases. Squamous cell carcinomas (left), show a negative correlation with BM, while adenocarcinomas (right) display the opposite behavior.

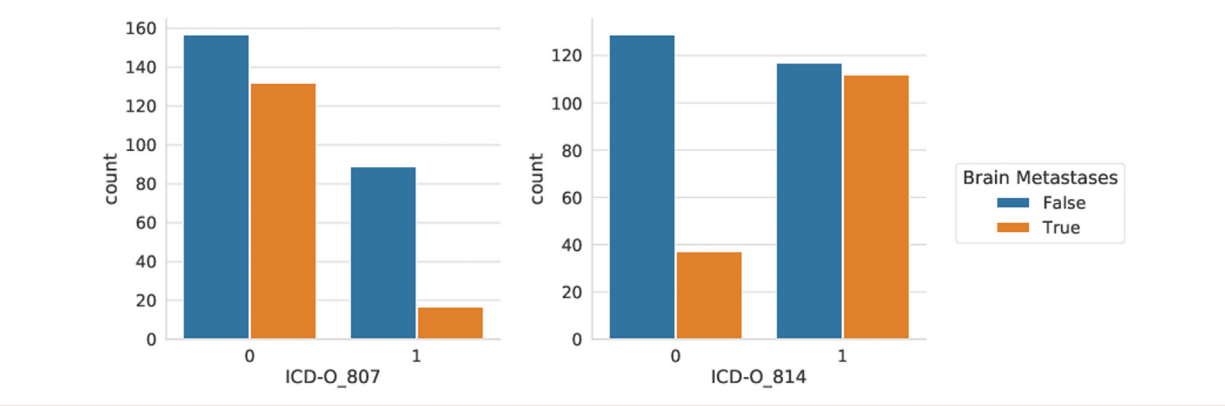


Table 4 Classification Metrics Evaluated for the Different Sets of Features

Data	Model	Sensitivity	Specificity	AUPRC	MCC
Clinical	Autogluon	6.16E-01	0.611	0.476	0.219
	Lreg	7.34E-01	0.572	**0.513**	0.293
	LightGBM	0.646	**0.647**	**0.513**	0.28
	RF	**0.741**	0.567	0.511	**0.295**
Radiomics	Autogluon	0.176	0.814	0.163	-0.009
	LReg	**0.514**	0.48	0.145	-0.004
	LightGBM	0.052	**0.935**	0.167	-0.012
	RF	0.236	0.776	**0.179**	**0.010**
Combined	Autogluon	0.219	**0.802**	0.162	0.019
	LReg	**0.467**	0.614	0.179	**0.061**
	LightGBM	0.232	0.801	0.191	0.032
	RF	0.267	0.788	**0.208**	0.042

The area under the precision-recall curve (AUPRC) was calculated instead of the area under the receiver operating characteristic (AUROC) as it is a critical metric in problems where properly classifying the positives is important.
The bold values indicate the best average value for the relevant metric among the models tested on the same set of features.

Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002

Figure 3 Flow chart showing number of patients enrolled in retrospective study cohort, excluded and analyzed.

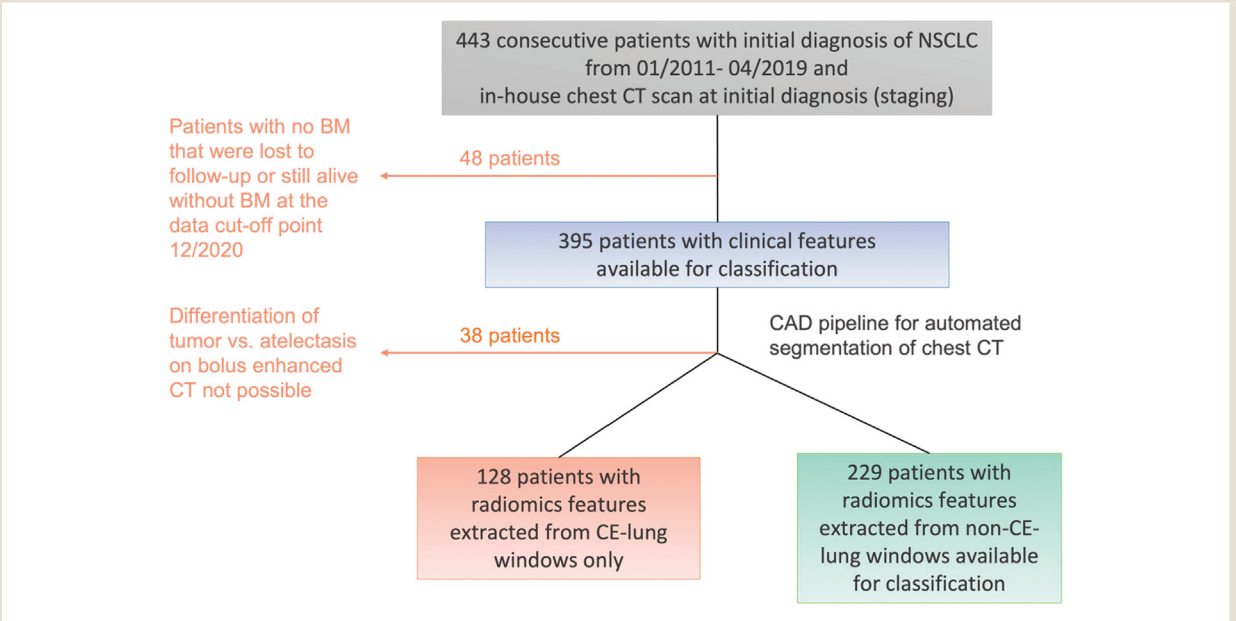
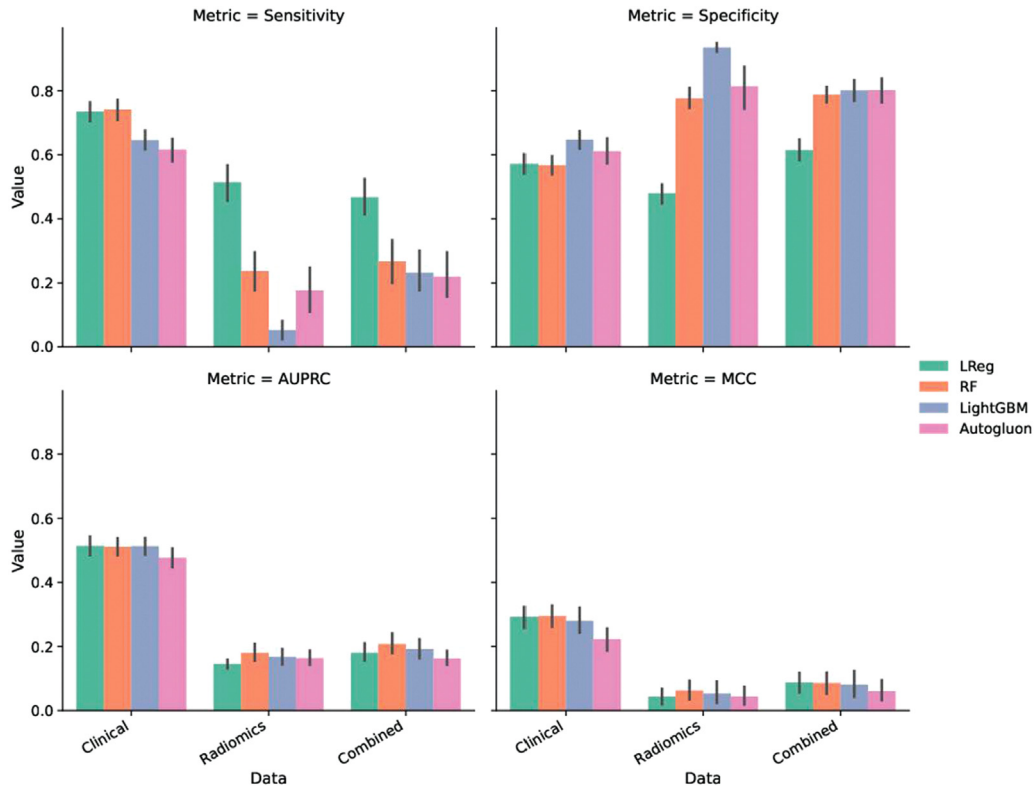


Figure 4 Classification result for different models and different input features.



Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002

the combination of clinical and radiomics features for this group achieved performance comparable to the use of the full clinical dataset. Resampling methods such as Synthetic Minority Oversampling Technique (SMOTE)[47] were tested on the set of radiomics features, but brought no significant improvement.

Importance of Clinical Features

Explaining the predictions of machine learning models is crucial to adopt such classifiers in clinical settings.[48] Despite the limited performance that classifiers have shown using clinical features as input, we aimed to analyze which features influence the predicted label the most. To this end, we employed 2 feature importance measures to dissect the workings of the RF classifier on the KNN5-imputed dataset. The RF model was selected for this analysis because it displayed sufficiently robust performance over the randomized train/test splits (Supplemental Figure S3). Within the same setting of nested cross-validation presented for the analysis of classification performance, we evaluated Gini importance, and SHAP values for the best model trained in each split. Based on these importance measures, we ranked the features in order of importance and analyzed how consistent these rankings are over the randomized seeds. The results are summarized in Figure 5. The 3 features that had shown significant correlation with BM (age, ICD-O code 807 = squamous cell carcinoma histology, and ICD-O code 814 = adenocarcinoma histology) consistently showed up as the highest ranking in both evaluation methods. The Gini Importance assigns higher rank to the patient's age, consistently with the fact that GI is known to be biased towards continuous variables or variables with many categories.[49] Examining the correlation between SHAP values and feature values offers insights into the trends that links certain features with the BM label (Supplemental Figure S4). Age and ICD-O code 807 displayed a negative correlation with the presence of BM, while ICD-O code 814 had a positive correlation, consistent with the results of the Kendall's Tau tests. In addition, the histological tumor grade and the presence of regional lymph node invasion at initial diagnosis appeared to be positively correlated[4] with a prediction of BM.

Identification of Patient Subpopulations

Among the first candidate subgroups, the algorithm identified the subgroups of female patients with adenocarcinoma histology (ICD-O code 814), and the subgroup of adenocarcinoma patients with no other distant metastases than BM. In both combinations, the incidence of BM was noticeably higher in the selected subgroups compared to the remaining population.

Discussion

In this work, we presented an analysis of the use of supervised machine learning models for predicting the development of BM in NSCLC patients using clinical and radiomics features.

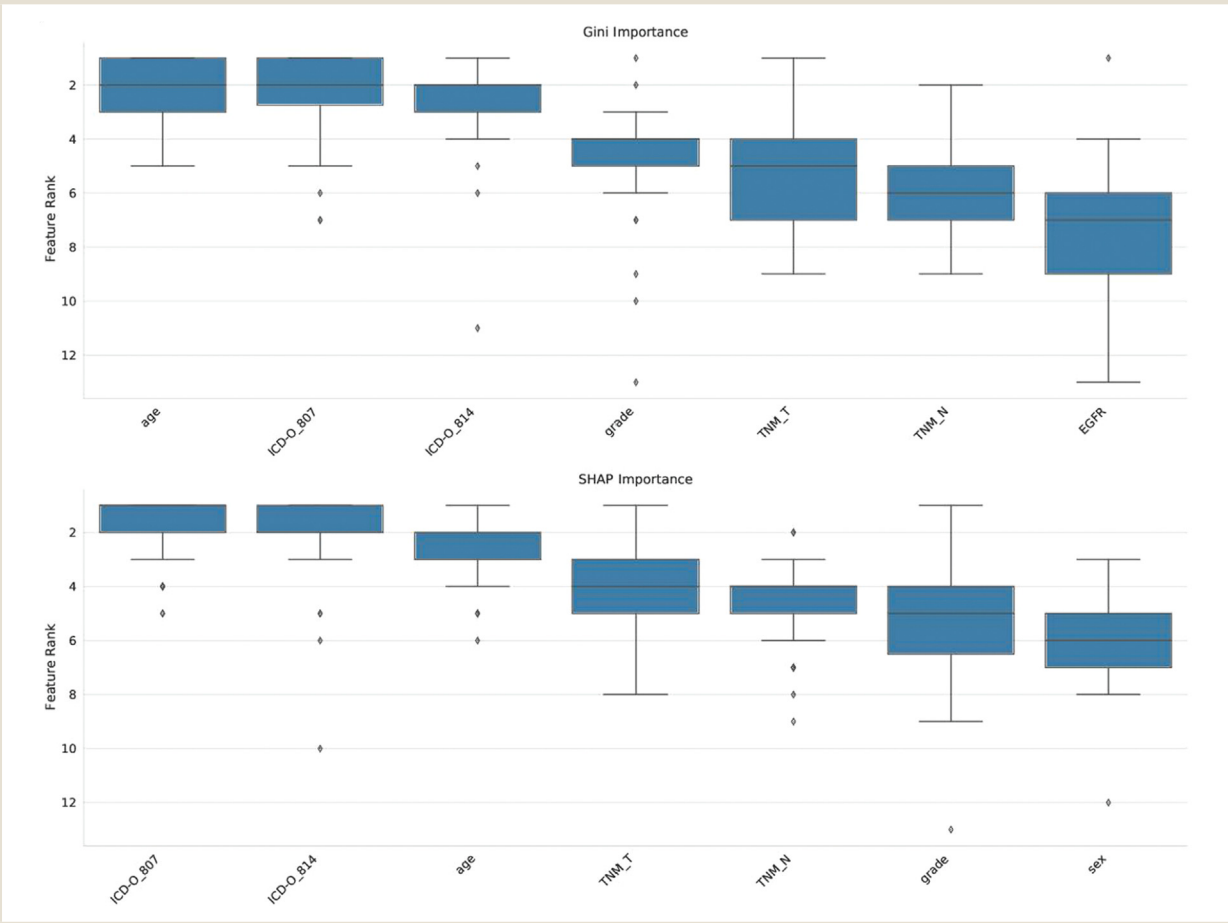
In our cohort, classification models that employed only clinical features as input showed the best performance when tuned to maximize recall to improve prospective studies, obtaining at best a sensitivity of $\sim$ 70% and a specificity just above 60%. Radiomics features performed suboptimally; however, this result is likely due to the limited number of samples for this dataset due to the heterogeneity of the existing imaging data (229 samples vs. 395 in the clinical dataset; Figure 3). In all experiments performed, the dataset size turned out to be a crucial element for the predictive performance of the models. Therefore, a robust assessment of the significance of radiomics features for predicting the development of BM in NSCLC patients will require further data collection. The models trained were optimized to achieve high sensitivity, as that would be the most valuable performance metrics when the model is employed for prospective studies, where a higher retrieval of true positives might be more valuable than reducing the false positives. Alternative clinical applications would require much a more careful balance of the performance metrics.

The analysis of the importance of the input features shows high agreement with the established literature, suggesting that the models are learning the relevant relationships between the clinical features and the development of BM. Age and ICD-O code 807 (squamous cell carcinoma histology) displayed a negative correlation with the presence of BM, while ICD-O code 814 (adenocarcinoma histology) had a positive correlation. The histological tumor grade and the presence of regional lymph node invasion at initial diagnosis appeared to be positively correlated with a prediction of BM.[4,34,50-54] It is still not fully explored why younger patients have a higher BM risk. Several studies indicate that younger patients harbor more targetable driver mutations compared to older patients, including a higher rate of EGFR and ALK gene alterations. The difference in prevalence of targetable mutations combined with a better performance status in younger patients may lead to a longer overall survival - long enough to experience intracranial spread.[55,56] There are other hypotheses, including differences in the angiogenic microenvironment[57] or expression of biomarkers, such as vascular endothelial growth factor, Ki-67, and caspase-3 that might create a more favorable environment for seeding.[58] In our analysis, female patients appeared more prone to developing BM, which is reflected in the correlation of SHAP values with the feature. This is potentially an indirect effect, since women are more likely to develop adenocarcinomas,[59,60] and female sex is a predictive factor for EGFR mutations,[33,35] which are themselves positively correlated with the development of BM.[33,61] Overall, the analysis suggests that these highly ranked features should be included in any machine learning model that aims to robustly predict the development of BM. Some known predictors, such as the presence of EGFR mutations, did not show statistically significant results, possibly due to the small size of the dataset (only 12% of patients tested showed a mutation) that does not offer a high enough signal-to-noise ratio to detect these features as significant, compounded with the high fraction of missing values (mutation analysis was not performed in 43% of patients). In addition, the rate in the primarily Caucasian population is expected to be very low even in complete datasets.[62,63]

Among the first candidate subgroups, the algorithm identified the subgroups of female patients with adenocarcinoma histology (ICD-O code 814), and the subgroup of adenocarcinoma patients with no other distant metastases. In both combinations, the incidence of BM was noticeably higher in the selected subgroups compared to the remaining population. However, it is unclear whether their significance is due to real clinical factors or simply due to selec-

Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002

Figure 5 Ranking of features according to 2 different evaluation methods over 30 randomized seeds using the RF model of the previous section. ICD-O code 807 = squamous cell carcinoma histology; ICD-O code 814 = adenocarcinoma histology; Table 2 describes the individual features.



tion biases. In the first case, for example, we already noted how female patients are likelier to develop tumors with adenocarcinoma histology,[59] which is directly correlated with the development of BM. Additional studies with more data will be required to make robust conclusions on the clinical relevance of these subpopulations of patients.

Next to the sample size, another important limitation of our study collective is the fact that systemic treatment and testing for driver mutations has completely changed over the recruitment period. Therefore, the true number of patients harboring driver mutations and their possible impact on the population remains unclear. In addition, the introduction of checkpoint inhibitors in the middle of the recruitment period might have had an impact on overall survival,[64] which, as aforementioned, could leave sufficient time for intracranial spread. Information on anticancer therapies was not documented in a structured manner and was therefore not available for analysis.

Recently, 956 stage II to IV NSCLC patients from the AACR Project GENIE Biopharma Consortium dataset were analyzed for the prediction of BM. In addition to the variables we considered, information on race, anticancer therapies and next-generation sequencing (NGS) features were available. With this broader information, an area under the receiver operating characteristic (AUROC) of up to 0.72 could be achieved on the test set. However, using an RF model, demographics alone performed as well as demographics, medication and genetic analyses combined. Univariate features associated with BM in this study were treatment with etoposide, Asian race, presence of bone metastases at NSCLC diagnosis, mutations in TP53 and EGFR, amplifications of ERBB2 and EGFR, and deletions of RB1, CDKN2A and CDKN2B. Univariate features inversely associated with BM were older age, treatment with nivolumab, vinorelbine, alectinib, pembrolizumab, atezolizumab, and gemcitabine, as well as mutations in NOTCH1, and KRAS.[65]

Another promising approach in the identification of high-risk subgroups is the combination of clinical parameters with gene expression signatures (RNA sequencing data) found in primary tumors, that are linked specifically to BM and not associated with

Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002

the development of metastasis to other sites or simply to recurrence of disease. Particularly, the oxidative phosphorylation pathway seems to be strongly associated with the risk of BM.[66]

Limitations With Regard to Radiomics

Radiomics features could ultimately be obtained only from a relatively small sample size (229 samples) and therefore performed suboptimally. There are 2 reasons for this: First of all, central bronchogenic tumors can cause post-obstructive pneumonia resulting in lung volume loss that can induce atelectasis.[67] Bolus-enhanced CT performs poorly in the differentiation of tumor and atelectasis.[68] As a consequence, patients had to be excluded from the radiomic analysis in case we were unable to distinguish tumor tissue from nontumor tissue on bolus-enhanced CT even with simultaneous consideration of all additional information available from PET/CT images or MRI. The second reason lies in the challenge of radiomics feature stability analyzing real world data. While the overall goal for radiomics studies is to maintain homogeneity in hardware and image scan protocols, reality of long-term studies presents its performers with a dilemma: they have to accept that scanners from different manufacturers are used at different institutions and that hardware will be upgraded. Imaging for this study was conducted on 6 different scanners with consecutively varying acquisition parameters. In general, radiomic features are called reproducible if they remain the same when extracted using different equipment, different image acquisition settings, or different post processing filters.[69] While phantom studies indicate that variable x-ray tube currents are unlikely to have a large effect on radiomic features extracted from CT images of textured objects such as tumors,[70] convolution kernel differences seem to have a large impact on radiomic feature robustness with poor stability of intensity, texture, and wavelet features.[71,72] Comparing contrast with noncontrast enhanced CT in multicentric patient data sets, a high impact on robustness of radiomic features was shown.[71,73] IV contrast influence can be controlled by using a selection of contrast enhanced or noncontrast enhanced images only. However, as Kakino et al.[72] report, some of the radiomic features in contrast-enhanced images are affected by the substrate itself and largely depend on patient characteristics such as patient weight, age, tumor volume or performance status. Even though all images in our study were iteratively reconstructed with a sharp kernel (b60) and displayed in standard lung window setting, they were either reconstructed from IV contrast enhanced raw data or from noncontrast enhanced data and stored in the picture archiving and communication system (PACS). As common at most hospitals, raw data was stored strictly local on scanners and deleted after 1 to 2 weeks. The majority of patients with images reconstructed from IV contrast enhanced raw data had been examined at the institute of Neuroradiology where they were admitted with BM. As a consequence, the assignment of IV contrast was highly correlated to the presence of BM (p-value for χ^2 test of independence: $2 * 10\text{-}32$), the inclusion of both IV positive and IV negative samples would have led to information leakage in the classification task. As mentioned, resampling methods such as SMOTE were tested on the set of radiomics features, but brought no significant improvement. These observations suggest that, for any future studies where similar classification tasks are examined, the collection of a higher number of samples is to be prioritized.

Conclusion

In this work, we presented an analysis of the use of supervised machine learning models for predicting the development of BM in NSCLC patients. Analysis of the importance of the input features showed a high agreement with the established literature, suggesting that the prediction models are learning the relevant relationships between the clinical features and the development of BM. To improve the performance of the models, the collection of a higher number of samples is to be prioritized. Employed prospectively at initial diagnosis, such models can help identify high-risk patients who might benefit from surveillance brain MRI for early detection and treatment of newly developed, asymptomatic BM. Cost-effectiveness of a surveillance program for high-risk subgroups would have to be investigated.

Clinical Practice Points

- Non-small Cell Lung Cancer (NSCLC) is among the cancer subtypes with the highest incidence of brain metastases (BM), which are associated with increased mortality.
- Early detection of BM is crucial to improve clinical prospects of NSCLC patients. While brain MRI is recommended at initial diagnosis and in case of neurological symptoms during follow-up, guidelines do not advise routine surveillance brain MRI.
- We trained and validated classifier models on clinical and/or 3D-radiomics-features of the primary tumor from staging chest-CT to identify patients with a high risk of developing BM.
- Predictive models based on clinical features offered the best performance (tuned to maximize recall: sensitivity$\sim$ 70%, specificity $\sim$ 60%). Radiomics features failed to provide sufficient information, likely due to the heterogeneity of imaging data.
- Analysis of the importance of the input features showed a high agreement with the established literature, suggesting that the prediction models are learning the relevant relationships between the clinical features and the development of BM.
- To improve the performance of the models, the collection of a higher number of samples is to be prioritized. Employed prospectively at initial diagnosis, such models can help identify high-risk patients who might benefit from surveillance brain MRI for early detection and treatment of newly developed, asymptomatic BM.

Disclosure

M.S. reports research funding to institution (but not for this specific project) from Astra Zeneca, honoraria for advisory boards from Astra Zeneca, BMS, MSD/Merck, Roche, Pfizer, Novartis, Boehringer Ingelheim, Takeda, Abbvie, Johnson & Johnson, Amgen, Tesaro, Eli Lilly, honoraria for educational lectures from Astra Zeneca, BMS, Novartis, Pfizer, Boehringer Ingelheim, Amgen and Eli Lilly. The authors have no conflicts of interest to report with regard to this project.

Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002

CRediT authorship contribution statement

Giovanni Visonà: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Lisa M. Spiller:** Data curation, Formal analysis, Investigation, Writing – review & editing. **Sophia Hahn:** Data curation, Formal analysis, Investigation, Writing – review & editing. **Elke Hattingen:** Resources, Writing – review & editing. **Thomas J. Vogl:** Resources, Writing – review & editing. **Gabriele Schweikert:** Writing – review & editing. **Katrin Bankov:** Data curation, Writing – review & editing. **Melanie Demes:** Data curation, Writing – review & editing. **Henning Reis:** Data curation, Writing – review & editing. **Peter Wild:** Resources, Writing – review & editing. **Pia S. Zeiner:** Validation, Writing – review & editing. **Fabian Acker:** Validation, Writing – review & editing. **Martin Sebastian:** Resources, Validation, Writing – review & editing. **Katharina J. Wenger:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Availability of Data and Material

The data that support the findings of this study are available from the corresponding author, KW, upon reasonable request.

Consent to Participate

Written informed consent was waived by the institutional review board in this retrospective study.

Consent for Publication

Written consent for publication was waived by the by the institutional review board in this retrospective study.

Ethics Approval

This study was approved by the ethical committee of University Hospital Frankfurt (UCT-3-2018, Geschäfts-Nummer: 4/09) and was in accordance with the 1964 Helsinki Declaration and its later amendments.

Acknowledgments

This work was supported by the Frankfurter Forschungs Förderung (FFF) of the University Hospital Frankfurt. In addition, it has received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Marie Skłodowska-Curie Grant Agreement No. 813533-MSCA-ITN-2018.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.cllc.2023.08.002.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J Clin*. 2021;71(3):209–249. doi:10.3322/caac.21660.
2. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of lung tumors. *J Thoracic Oncol*. 2015;10(9):1243–1260. doi:10.1097/JTO.0000000000000630.
3. Barnholtz-Sloan JS, Sloan AE, Davis FG, Vigneau FD, Lai P, Sawaya RE. Incidence proportions of brain metastases in patients diagnosed (1973 to 2001) in the metropolitan detroit cancer surveillance system. *JCO*. 2004;22(14):2865–2872. doi:10.1200/JCO.2004.12.149.
4. Mujoomdar A, Austin JHM, Malhotra R, et al. Clinical predictors of metastatic disease to the brain from non-small cell lung carcinoma: primary tumor size, cell type, and lymph node metastases. *Radiology*. 2007;242(3):882–888. doi:10.1148/radiol.2423051707.
5. Schouten LJ, Rutten J, Huveneers HAM, Twijnstra A. Incidence of brain metastases in a cohort of patients with carcinoma of the breast, colon, kidney, and lung and melanoma. *Cancer*. 2002;94(10):2698–2705. doi:10.1002/cncr.10541.
6. Iuchi T, Shingyoji M, Itakura M, et al. Frequency of brain metastases in non-small-cell lung cancer, and their association with epidermal growth factor receptor mutations. *Int J Clin Oncol*. 2015;20(4):674–679. doi:10.1007/s10147-014-0760-9.
7. Cagney DN, Martin AM, Catalano PJ, et al. Incidence and prognosis of patients with brain metastases at diagnosis of systemic malignancy: a population-based study. *Neuro Oncol*. 2017;19(11):1511–1521. doi:10.1093/neuonc/nox077.
8. Ge M, Zhuang Y, Zhou X, Huang R, Liang X, Zhan Q. High probability and frequency of EGFR mutations in non-small cell lung cancer with brain metastases. *J Neurooncol*. 2017;135(2):413–418. doi:10.1007/s11060-017-2590-x.
9. Hsu F, De Caluwe A, Anderson D, Nichol A, Toriumi T, Ho C. EGFR mutation status on brain metastases from non-small cell lung cancer. *Lung Cancer*. 2016;96:101–107. doi:10.1016/j.lungcan.2016.04.004.
10. Zhang I, Zaorsky NG, Palmer JD, Mehra R, Lu B. Targeting brain metastases in ALK-rearranged non-small-cell lung cancer. *Lancet Oncol*. 2015;16(13):e510–e521. doi:10.1016/S1470-2045(15)00013-3.
11. Rangachari D, Yamaguchi N, VanderLaan PA, et al. Brain metastases in patients with EGFR-mutated or ALK-rearranged non-small-cell lung cancers. *Lung Cancer*. 2015;88(1):108–111. doi:10.1016/j.lungcan.2015.01.020.
12. Gainor JF, Tseng D, Yoda S, et al. Patterns of metastatic spread and mechanisms of resistance to crizotinib in ROS1-positive non-small-cell lung cancer. *JCO Precis Oncol*. 2017;1 PO.17.00063. doi:10.1200/PO.17.00063.
13. Patil T, Smith DE, Bunn PA, et al. The incidence of brain metastases in stage IV ROS1-rearranged non–small cell lung cancer and rate of central nervous system progression on crizotinib. *J Thoracic Oncol*. 2018;13(11):1717–1726. doi:10.1016/j.jtho.2018.07.001.
14. Sperduto PW, Yang TJ, Beal K, et al. Estimating survival in patients with lung cancer and brain metastases: an update of the graded prognostic assessment for lung cancer using molecular markers (Lung-molGPA). *JAMA Oncol*. 2017;3(6):827–831. doi:10.1001/jamaoncol.2016.3834.
15. Planchard D, Popat S, Kerr K, et al. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2018;29(Suppl 4):iv192–iv237. doi:10.1093/annonc/mdy275.
16. Kuhn MJ, Hammer GM, Swenson LC, Youssef HT, Gleason TJ. MRI evaluation of "solitary" brain metastases with triple-dose gadoteridol: comparison with contrast-enhanced CT and conventional-dose gadopentetate dimeglumine MRI studies in the same patients. *Comput Med Imaging Graph*. 1994;18(5):391–399. doi:10.1016/0895-6111(94)90011-6.
17. de Groot PM, Chung JH, et al.Expert Panel on Thoracic Imaging ACR appropriateness Criteria® noninvasive clinical staging of primary lung cancer. *J Am Coll Radiol*. 2019;16(5S):S184–S195. doi:10.1016/j.jacr.2019.02.008.
18. Schneider BJ, Ismaila N, Aerts J, et al. Lung cancer surveillance after definitive curative-intent therapy: ASCO guideline. *JCO*. 2020;38(7):753–766. doi:10.1200/JCO.19.02748.
19. Gavrilovic IT, Posner JB. Brain metastases: epidemiology and pathophysiology. *J Neurooncol*. 2005;75(1):5–14. doi:10.1007/s11060-004-8093-6.
20. Jacobs L, Kinkel WR, Vincent RG. Silent" brain metastasis from lung carcinoma determined by computerized tomography. *Arch Neurol*. 1977;34(11):690–693. doi:10.1001/archneur.1977.00500230060010.
21. Yokoi K, Kamiya N, Matsuguma H, et al. Detection of brain metastasis in potentially operable non-small cell lung cancer: a comparison of CT and MRI. *Chest*. 1999;115(3):714–719. doi:10.1378/chest.115.3.714.
22. Besse B, Le Moulec S, Mazières J, et al. Bevacizumab in patients with nonsquamous non–small cell lung cancer and asymptomatic, untreated brain metastases (BRAIN): a nonrandomized, phase II study. *Clin Cancer Res*. 2015;21(8):1896–1903. doi:10.1158/1078-0432.CCR-14-2082.
23. Page S, Milner-Watts C, Perna M, et al. Systemic treatment of brain metastases in non-small cell lung cancer. *Eur J Cancer*. 2020;132:187–198. doi:10.1016/j.ejca.2020.03.006.
24. Jena A, Taneja S, Talwar V, Sharma JB. Magnetic resonance (MR) patterns of brain metastasis in lung cancer patients: correlation of imaging findings with symptom. *J Thoracic Oncol*. 2008;3(2):140–144. doi:10.1097/JTO.0b013e318161d775.
25. Kim SY, Kim JS, Park HS, et al. Screening of brain metastasis with limited magnetic resonance imaging (MRI): clinical implications of using limited brain

Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002

MRI during initial staging for non-small cell lung cancer patients. *J Korean Med Sci*. 2005;20(1):121. doi:10.3346/jkms.2005.20.1.121.
26. Kim M, Suh CH, Lee SM, et al. Development of brain metastases in patients with non-small cell lung cancer and no brain metastases at initial staging evaluation: cumulative incidence and risk factor analysis. *AJR Am J Roentgenol*. 2021;217(5):1184–1193. doi:10.2214/AJR.21.25787.
27. Pugh TJ, Gaspar LE. Prophylactic cranial irradiation for patients with lung cancer. *Clin Lung Cancer*. 2007;8(6):365–368. doi:10.3816/CLC.2007.n.016.
28. De Ruysscher D, Dingemans AMC, Praag J, et al. Prophylactic cranial irradiation versus observation in radically treated stage III non-small-cell lung cancer: a randomized phase III NVALT-11/DLCRG-02 study. *J Clin Oncol*. 2018;36(23):2366–2377. doi:10.1200/JCO.2017.77.5817.
29. Gore EM, Bae K, Wong SJ, et al. Phase III comparison of prophylactic cranial irradiation versus observation in patients with locally advanced non-small-cell lung cancer: primary analysis of radiation therapy oncology group study RTOG 0214. *J Clin Oncol*. 2011;29(3):272–278. doi:10.1200/JCO.2010.29.1609.
30. Hirsch FR, Suda K, Wiens J, Bunn PA. New and emerging targeted treatments in advanced non-small-cell lung cancer. *Lancet*. 2016;388(10048):1012–1024. doi:10.1016/S0140-6736(16)31473-8.
31. Reck M, Rabe KF. Precision diagnosis and treatment for advanced non-small-cell lung cancer. *N Engl J Med*. 2017;377(9):849–861. doi:10.1056/NEJMra1703413.
32. Shi L, He Y, Yuan Z, et al. Radiomics for response and outcome assessment for non-small cell lung cancer. *Technol Cancer Res Treat*. 2018;17. doi:10.1177/1533033818782788.
33. Burel-Vandenbos F, Ambrosetti D, Coutts M, Pedeutour F. EGFR mutation status in brain metastases of non-small cell lung carcinoma. *J Neurooncol*. 2013;111(1):1–10. doi:10.1007/s11060-012-0990-5.
34. Waqar SN, Waqar SH, Trinkaus K, et al. Brain metastases at presentation in patients with non–small cell lung cancer. *Am J Clin Oncol*. 2018;41(1):36–40. doi:10.1097/COC.0000000000000230.
35. Yawn BP, Wollan PC, Schroeder C, Gazzuola L, Mehta M. Temporal and gender-related trends in brain metastases from lung and breast cancer. *Minn Med*. 2003;86(12):32–37.
36. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc: Series B (Methodol)*. 1995;57(1):289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.
37. Puoliväli T, Palva S, Palva JM. Influence of multiple hypothesis testing on reproducibility in neuroimaging research: a simulation study and Python-based software. *J Neurosci Methods*. 2020;337. doi:10.1016/j.jneumeth.2020.108654.
38. Ke G, Meng Q, Finley T, et al. *Advances in Neural Information Processing Systems*. LightGBM: a highly efficient gradient boosting decision tree. Long Beach, CA, USA: Curran Associates, Inc.; 2017 Vol 30. https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf, Accessed November 4, 2022.
39. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104–e107. doi:10.1158/0008-5472.CAN-17-0339.
40. Fortin JP, Cullen N, Sheline YI, et al. Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage*. 2018;167:104–120. doi:10.1016/j.neuroimage.2017.11.024.
41. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5–32. doi:10.1023/A:1010933404324.
42. Erickson N, Mueller J, Shirkov A, et al. AutoGluon-Tabular: robust and accurate AutoML for structured data. Accessed November 4, 2022. http://arxiv.org/abs/2003.06505
43. Cawley G, Talbot N. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Machine Learning Res*. 2010;11:2079–2107. http://jmlr.csail.mit.edu/papers/v11/cawley10a.html. Accessed November 4, 2022.
44. Atzmueller M. Subgroup discovery: subgroup discovery. *WIREs Data Mining Knowl Discov.*. 2015;5(1):35–49. doi:10.1002/widm.1144.
45. Helal S. subgroup discovery algorithms: a survey and empirical evaluation. *J Comput Sci Technol*. 2016;31(3):561–576. doi:10.1007/s11390-016-1647-1.
46. Lemmerich F, Becker M, et al. Pysubgroup: easy-to-use subgroup discovery in Python. In: Brefeld U, Curry E, Daly E, et al., eds. *Machine Learning and Knowledge Discovery in Databases* Springer International Publishing; 2019:658–662. doi:10.1007/978-3-030-10997-4_46.
47. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. 2011. doi:10.48550/ARXIV.1106.1813
48. Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. *IEEE Trans Neural Netw Learning Syst*. 2021;32(11):4793–4813. doi:10.1109/TNNLS.2020.3027314.
49. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinform*. 2007;8(1):25. doi:10.1186/1471-2105-8-25.
50. Waqar SN, Samson PP, Robinson CG, et al. Non-small-cell lung cancer with brain metastasis at presentation. *Clin Lung Cancer*. 2018;19(4):e373–e379. doi:10.1016/j.cllc.2018.01.007.
51. Ceresoli GL, Reni M, Chiesa G, et al. Brain metastases in locally advanced nonsmall cell lung carcinoma after multimodality treatment: risk factors analysis. *Cancer*. 2002;95(3):605–612. doi:10.1002/cncr.10687.
52. Wang S-Y, X Ye, Ou W, Lin Y-B, Zhang B-B, Yang H. Risk of cerebral metastases for postoperative locally advanced non-small-cell lung cancer. *Lung Cancer*. 2009;64(2):238–243. doi:10.1016/j.lungcan.2008.08.012.
53. Won YW, Joo J, Yun T, et al. A nomogram to predict brain metastasis as the first relapse in curatively resected non-small cell lung cancer patients. *Lung Cancer*. 2015;88(2):201–207. doi:10.1016/j.lungcan.2015.02.006.
54. Lee H, Jeong SH, Jeong BH, et al. Incidence of brain metastasis at the initial diagnosis of lung squamous cell carcinoma on the basis of stage, excluding brain metastasis. *J Thoracic Oncol*. 2016;11(3):426–431. doi:10.1016/j.jtho.2015.11.007.
55. Subramanian J, Morgensztern D, Goodgame B, et al. Distinctive characteristics of non-small cell lung cancer (NSCLC) in the young: a surveillance, epidemiology, and end results (SEER) analysis. *J Thoracic Oncol*. 2010;5(1):23–28. doi:10.1097/JTO.0b013e3181c41e8d.
56. Lara MS, Brunson A, Wun T, et al. Predictors of survival for younger patients less than 50 years of age with non-small cell lung cancer (NSCLC): a California Cancer Registry analysis. *Lung Cancer*. 2014;85(2):264–269. doi:10.1016/j.lungcan.2014.04.007.
57. Fidler IJ, Yano S, Zhang RD, Fujimaki T, Bucana CD. The seed and soil hypothesis: vascularisation and brain metastases. *Lancet Oncol*. 2002;3(1):53–57. doi:10.1016/s1470-2045(01)00622-2.
58. Saad AG, Yeap BY, Thunnissen FBJM, et al. Immunohistochemical markers associated with brain metastases in patients with nonsmall cell lung carcinoma. *Cancer*. 2008;113(8):2129–2138. doi:10.1002/cncr.23826.
59. Barrera-Rodriguez R, Morales-Fuentes J. Lung cancer in women. *Lung Cancer (Auckl)*. 2012;3:79–89. doi:10.2147/LCTT.S37319.
60. North CM, Christiani DC. Women and lung cancer: what is new? *Semin Thorac Cardiovasc Surg*. 2013;25(2):87–94. doi:10.1053/j.semtcvs.2013.05.002.
61. Shin DY, Na II, Kim CH, Park S, Baek H, Yang SH. EGFR mutation and brain metastasis in pulmonary adenocarcinomas. *J Thorac Oncol*. 2014;9(2):195–199. doi:10.1097/JTO.0000000000000069.
62. Cortot AB, Italiano A, Burel-Vandenbos F, Martel-Planche G, Hainaut P. KRAS mutation status in primary nonsmall cell lung cancer and matched metastases. *Cancer*. 2010;116(11):2682–2687. doi:10.1002/cncr.25014.
63. Daniele L, Cassoni P, Bacillo E, et al. Epidermal growth factor receptor gene in primary tumor and metastatic sites from non-small cell lung cancer. *J Thorac Oncol*. 2009;4(6):684–688. doi:10.1097/JTO.0b013e3181a52359.
64. Neumann M, Murphy N, Seetharamu N. The evolving role of PD-L1 inhibition in non-small cell lung cancer: a review of durvalumab and avelumab. *Cancer Med J*. 2022;5(1):31–45.
65. Rahman P, LeNoue-Newton M, Chaugai S, et al. Clinical and genomic predictors of brain metastases (BM) in non-small cell lung cancer (NSCLC): An AACR Project GENIE analysis. *JCO*. 2021;39(15_suppl) 2032-2032. doi:10.1200/JCO.2021.39.15_suppl.2032.
66. Kamer I, Steuerman Y, Daniel-Meshulam I, et al. Predicting brain metastasis in early stage non-small cell lung cancer patients by gene expression profiling. *Transl Lung Cancer Res*. 2020;9(3). https://tlcr.amegroups.com/article/view/40177. Accessed November 11, 2022.
67. Yang RM, Li L, Wei XH, et al. Differentiation of central lung cancer from atelectasis: comparison of diffusion-weighted MRI with PET/CT. *PLoS One*. 2013;8(4):e60279. doi:10.1371/journal.pone.0060279.
68. Qi LP, Zhang XP, Tang L, Li J, Sun YS, Zhu GY. Using diffusion-weighted MR imaging for tumor detection in the collapsed lung: a preliminary study. *Eur Radiol*. 2009;19(2):333–341. doi:10.1007/s00330-008-1134-3.
69. O'Connor JPB, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol*. 2017;14(3):169–186. doi:10.1038/nrclinonc.2016.162.
70. Mackin D, Ger R, Dodge C, et al. Effect of tube current on computed tomography radiomic features. *Sci Rep*. 2018;8(1):2354. doi:10.1038/s41598-018-20713-6.
71. Vuong D, Bogowicz M, Denzler S, et al. Comparison of robust to standardized CT radiomics models to predict overall survival for non-small cell lung cancer patients. *Med Phys*. 2020;47(9):4045–4053. doi:10.1002/mp.14224.
72. Kakino R, Nakamura M, Mitsuyoshi T, et al. Comparison of radiomic features in diagnostic CT images with and without contrast enhancement in the delayed phase for NSCLC patients. *Physica Medica*. 2020;69:176–182. doi:10.1016/j.ejmp.2019.12.019.
73. Jha AK, Mithun S, Jaiswar V, et al. Repeatability and reproducibility study of radiomic features on a phantom and human cohort. *Sci Rep*. 2021;11(1):2055. doi:10.1038/s41598-021-81526-8.
74. Scheel AH, Dietel M, Heukamp LC, et al. [Predictive PD-L1 immunohistochemistry for non-small cell lung cancer : current state of the art and experiences of the first German harmonization study]. *Pathologe*. 2016;37(6):557–567. doi:10.1007/s00292-016-0189-1.

Please cite this article as: Giovanni Visonà et al, Machine-Learning-Aided Prediction of Brain Metastases Development in Non–Small-Cell Lung Cancers, Clinical Lung Cancer, https://doi.org/10.1016/j.cllc.2023.08.002