

Machine Learning to Improve Treatment Selection for NSCLC Patients Treated with Immunotherapy Using Real World and Translational Data

Arsela Prelaj (✉ arsela20@hotmail.it)

Umberto I Policlinico di Roma <https://orcid.org/0000-0002-3863-088X>

Mattia Boeri

Fondazione IRCCS Istituto Nazionale dei Tumori

Alessandro Robuschi

Politecnico di Milano Dipartimento di Elettrotecnica: Politecnico di Milano Dipartimento di Elettronica Informazione e Bioingegneria

Roberto Ferrara

Fondazione IRCCS Istituto Nazionale dei Tumori

Claudia Proto

Fondazione IRCCS Istituto Nazionale dei Tumori

Giuseppe Lo Russo

Fondazione IRCCS Istituto Nazionale dei Tumori

Giulia Galli

Fondazione IRCCS Istituto Nazionale dei Tumori

Alessandro De Toma

Fondazione IRCCS Istituto Nazionale dei Tumori

Marta Brambilla

Fondazione IRCCS Istituto Nazionale dei Tumori

Mario Occhipinti

Fondazione IRCCS Istituto Nazionale dei Tumori

Sara Manglaviti

Fondazione IRCCS Istituto Nazionale dei Tumori

Teresa Beninato

Fondazione IRCCS Istituto Nazionale dei Tumori

Achille Bottiglieri

Fondazione IRCCS Istituto Nazionale dei Tumori

Giacomo Massa

Fondazione IRCCS Istituto Nazionale dei Tumori

Emma Zattarin

Fondazione IRCCS Istituto Nazionale dei Tumori

Rosaria Gallucci

Fondazione IRCCS Istituto Nazionale dei Tumori

Edoardogregorio Galli

Fondazione IRCCS Istituto Nazionale dei Tumori

Monica Ganzinelli

Fondazione IRCCS Istituto Nazionale dei Tumori

Gabriella Sozzi

Fondazione IRCCS Istituto Nazionale dei Tumori

Filippo G. M. De Braud

Fondazione IRCCS Istituto Nazionale dei Tumori

Marina Chiara Garassino

Fondazione IRCCS Istituto Nazionale dei Tumori

Marcello Restelli

Politecnico di Milano Dipartimento di Bioingegneria: Politecnico di Milano Dipartimento di Elettronica
Informazione e Bioingegneria

Alessandra Laura Giulia Pedrocchi

Politecnico di Milano Dipartimento di Bioingegneria: Politecnico di Milano Dipartimento di Elettronica
Informazione e Bioingegneria

Francesco Trovò

Politecnico di Milano Dipartimento di Bioingegneria: Politecnico di Milano Dipartimento di Elettronica
Informazione e Bioingegneria

Research

Keywords: non-small cell lung cancer, immunotherapy, biomarker, artificial intelligence, machine learning.

Posted Date: September 14th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-850506/v1>

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

[Read Full License](#)

Abstract

Background

In advanced Non-Small Cell Lung Cancer (NSCLC), Programmed Death Ligand 1 (PD-L1) remains the only used biomarker to candidate patients to immunotherapy (IO) with many limits. Given the complex dynamics of the immune system it is improbable that a single biomarker could be able to profile prediction with high accuracy. A promising solution cope with this complexity is provided by Artificial Intelligence (AI) and Machine Learning (ML), which are techniques able to analyse and interpret big multifactorial data. The present study aims at using AI tools to improve response and efficacy prediction in NSCLC patients treated with IO.

Methods

Real world data (clinical data, PD-L1, histology, molecular, lab tests) and the blood microRNA signature classifier (MSC), which include 24 different microRNAs, were used. Patients were divided into responders (R), who obtained a complete or partial response or stable disease as best response, and non-responders (NR), who experienced progressive or hyperprogressive disease and those who died before the first radiologic evaluation. Moreover, we used the same data to determine if the overall survival of the patients was likely to be shorter or longer than 24 months from baseline IO. For A literature review and forward feature selection technique was used to extract a specific subset of the patients data. To develop the final predictive model, different ML methods have been tested, i.e., Feedforward Neural Network (FFNN), Logistic Regression (LR), K-nearest neighbors (K-NN), Support Vector Machines (SVM), and Random Forest (RF).

Results

200 patients were included. 164 out of 200 (i.e., only those patients with PD-L1 data available) were considered in the model, 73 (44.5%) were R and 91 (55.5%) NR. Overall, the best model was the LR and included 5 features: 2 clinical features including the ECOG performance status and IO-line of therapy; 1 tissue feature such as PD-L1 tumour expression; and 2 blood features including the MSC test and the neutrophil-to-lymphocyte ratio (NLR). The model predicting R/NR of the patient achieves accuracy ACC=0.756, F1 score F1=0.722, and Area Under the ROC Curve AUC=0.82. The use of the PD-L1 alone has an ACC=0.655. The accuracy of the ML models excluding some of the features from the model were as follow: without PD-L1 value (ACC=0.726), MSC (ACC=0.750), and both PD-L1 and MSC (ACC=0.707), i.e., considering only clinical features. At data cut-off (Nov 2020), median Overall Survival (mOS) for R was 38.5 months (m) (95%IC 23.9 - 53.1) vs 3.8 m (95%IC 2.8 - 4.7) for NR, with $p < 0.001$. LR was the most performing model in predicting patients with long survival (24-months OS), achieving ACC=0.839, F1=0.908, and AUC=0.87.

Conclusions

The results suggest that the integration of multifactorial data provided by ML techniques is a useful tool to improve personalized selection of NSCLC patients candidates to IO. In particular, compare to PD-L1 alone the expected improvement was around 10%. In particular, the model shows that the higher the ECOG, NLR value, IO-line, and MSC test level the lower the response, and the higher PD-L1 the higher the response. Considering the difference in survival among R and NR groups, these results suggest that the model can also be used to indirectly predict survival. Moreover, a second model was able to predict long survival patients with good accuracy.

Background

Lung cancer is the leading cancer-related death worldwide with around 470.000 new cases and 390.000 deaths in Europe. Non-Small Cell Lung Cancer (NSCLC) is the most common histology for around 85% (1). Until 2015 the median OS of patients with metastatic NSCLC was around 12 months (2). The advent of Immunotherapy (IO) has radically changed the treatment paradigm of many cancers including NSCLC prolonging survival of metastatic patients from 12 to a median of around 24 months (2). Some of them who responded better to IO reached longer survival up to or more than 5 years (3). However, only 30–50% of patients will benefit from IO in the long term (4–6).

Currently in clinical practice, Programmed Death-Ligand 1 (PD-L1) is the only used biomarker to predict response to IO. However, its predictive performance is not satisfactory (around 30–50%) (7). Beyond PD-L1, several other biomarkers have been identified and used to profile patient prediction, including Tumour Mutational Burden (TMB) (8), tumour microenvironment (TME) (9), microRNA (miRNA) (10), immune gene signatures (11), gut microbiome (12), radiomics (13), or baseline clinical features or their combination in different scores (14, 15).

Indeed, it is implausible that a single biomarker is able to profile prediction or prognosis with a high accuracy, since the immune system displays dynamics complexity when interacting with its TME. To handle with the density of the available data, Artificial Intelligence (AI) frameworks and, more specifically, Machine Learning (ML) techniques, provide efficient, pioneering, and theoretically sound approaches to construct decision-making tools providing individualized prediction (16).

Among molecular biomarkers, the plasma microRNA signature classifier (MSC), reflecting an immunosuppressive host status, was here considered (10). It was previously trained in lung cancer screening cohorts to evaluate the individual risk to develop the aggressive form of the disease (17;18). More recently, the MSC prognostic value was also validated in advanced NSCLC patients treated with single agent IO (19) and its combination with different clinical scores confirmed its independency from other prognostic features in this setting (20).

This study aims to integrate real word data and MSC test to develop a machine learning algorithm to predict response and efficacy to IO in NSCLC patients. The study also investigates the role of the MSC test and its added value to the algorithm prediction capability, being this latter test costly and still not included in the standard clinical practice as a predictive/prognostic biomarker.

Material And Methods

Study population

From July 2015 to Nov 2020, we conducted a prospective observational study (Apollo, INT 22_15) enrolling 200 consecutive aNSCLC patients receiving single-agent anti-PD-(L)-1 inhibitors in first- (n=70) or further-line therapy (n=130). Complete real-world data and whole blood samples were collected as per clinical practice. The MSC test was prospectively assessed in plasma samples collected at baseline IO.

Inclusion criteria were the followings: cytological/histological diagnosis of advanced NSCLC, patients (relapsed or stage IIIB to IV) that had received at least one infusion of single agent IO in first- or further-line. Patients without baseline IO MSC test information were excluded from the study.

This prospective study was conducted at Fondazione IRCCS Istituto Nazionale Tumori of Milan in Italy in collaboration with Politecnico di Milano, for the data analytics. This study was approved from the local ethical committee and all included patients signed informed consent prior plasma and data collection accomplished in accordance with the Declaration of Helsinki, Good Clinical Practice and local ethical guideline.

Real World Data Collection: clinical, blood, and tissue data

For this study, demographic, medical history, tumour stage, PD-L1, molecular and radiological data, concomitant medications, treatment responses and survival follow-up were collected and integrated to develop a new predictive model of response and efficacy to IO in NSCLC.

Omic Collection: MSC blood test

Whole blood was collected in 10 ml K2EDTA Vacutainer tubes and the plasma separated by two centrifugation steps. Total RNA was extracted from 200 µl plasma samples. MicroRNA expression was determined by quantitative reverse transcription PCR (RT-qPCR) as previously described (19,21).

The MSC algorithm using 24 miRNAs defined four different classes of risks: low (L) intermediate (I) and high (H) risk (18) and highly haemolysed (E). The fourth category E, thus not analysable plasma samples, due to the unspecific released of miRNAs in presence of blood cell lyses, was included (10) (Figure 1). Patients with this category were previously observed to have an intermediate prognosis between patients with H and I risk (20).

Treatment administration

IO was administered intravenously (IV) as monotherapy. Nivolumab was administered initially at a dose of 3 mg/kg and later, since May 2018 in Italy, at a fixed dose of 240 mg every 2 weeks (w). Pembrolizumab at a fixed dose of 200 mg as first line and at dose 2 mg/kg every 3w in second or third-line setting. Atezolizumab at a fixed dose of 1200 mg every 3w and durvalumab at a dose 10 mg/kg every

2w. Therapy was continued until progressive disease (PD), intolerable toxicity, withdrawal or death. Treatment beyond PD was allowed if there was a clinical benefit according to clinician's decision.

Radiological response evaluation

Baseline radiological evaluations included a baseline Total Body Computed Tomography (TB-CT) scan, subsequently performed every 3-4 cycles or every 9-12 weeks as per standard of care, or whenever progression was clinically suspected. Six categories of radiological response were taken into consideration in this study to assess tumour response. Four of them (standard categories) included in Response Evaluation Criteria in Solid Tumours (RECIST1.1) criteria: Complete Response (CR), Partial Response (PR), Stable Disease (SD), Progressive Disease (PD). Two additional categories were included: Hyper Progression Disease (HPD), an atypical pattern of response to single agent IO (an acceleration of the progression compared to the natural history of the disease) as defined by Ferrara et al. (22) and Lo Russo et al. (23). Eventually, not evaluable (NE) was the sixth category of those patients who died due to PD before the first radiological evaluation.

STATISTICS AND AI METHODOLOGY

Statistical analysis

164 patients out of the 200 patients included in the present study having available PD-L1 expression were used as dataset for the ML algorithms since the prediction given by PD-L1 is the unique biomarker used in clinical practice. Conversely, all the 200 patients were included in the survival analysis. The first endpoint of the study is prediction of responder (R) and non-responder (NR) patients. In the R group were included patients who obtained a CR, PR or SD as per RECIST 1.1 while in the NR group were included those patients who obtained a PD per RECIST1.1., or an HPD or NE response (as described above). Other endpoints were at 24-months Overall Survival (OS), and median progression-free survival (mPFS) and median OS (mOS). mOS is measured from the date of IO start therapy until death, or last follow-up. mPFS was calculated from the date of IO start until PD or death due to any cause, or last follow-up visit for alive patients without PD. Kaplan-Meier was used to calculate mPFS and mOS with their 95% confidence interval, and to generate survival curves. Cox's proportional hazards model were used to calculate the Hazard Ratio (HR) between R and NR groups according to OS and PFS.

Machine learning methods

After data collection, descriptive analysis and data processing were performed. A first step consisted in the selection of a set of 21 features which have been determined to be the most relevant ones based on the published literature on NSCLC patients treated with IO and clinician experience. Finally, in the case pair of features showed a linear correlation higher than 0.8, we removed one of them, as customary in ML studies. The result is the set of M=15 most relevant features, provided in Table 1.

The problem of predicting R and NR was modelled as a binary classification problem, where we want to learn an approximation of the real relationship between the i -th patient's feature vector and the response

, where a patient has 1 for NR, and 0 for R. The same modelling has been applied to the problem of estimating the survival at 24 months, i.e., a patient has 1 if the patient does not survive at least 24 months, and 0 if she/he does. Data corresponding to the 40 alive patients with less than 24 months were excluded from this second analysis.

We selected a set of appropriate techniques from the ML literature to perform the above-mentioned classification task. More specifically, we tested Feedforward Neural Network (FFNN), Logistic Regression (LR), K-nearest neighbours (K-NN), Support Vector Machines (SVM), and Random Forest (RF). We applied a feature selection approach to select the proper subset of the original M features appropriate for each method. More specifically, we used a forward feature selection using the AIC criterion as metric to select the most appropriate set of features for each method and the best method. The 5-fold cross-validation ACC and F1 scores for the analysed methods, as well as the leave one out AUC, with the corresponding 95% confidence intervals were computed using the bootstrap approach (in brackets). The implementation of the procedure has been performed in Matlab, and the code performing all the ML procedures is available at <https://trovo.faculty.polimi.it/downloads.html>.

Results

Patients' characteristics

Two hundred NSCLC patients treated with anti-PD-(L)-1 in first or further-line therapy were included in the survival analysis. Most patients were male (65%) and smokers (79.5%), median age was 67 years (range 60-74 years) and 38% of patients were older than 70 years. PD-L1 was $\geq 50\%$ in 53 (26.5%), 1%-49% in 59 (29.5%), <1% in 52 (26%) and unknown in 36 (18%) patients. Median ECOG-PS was 1 (range 0-1) with an ECOG PS 2 in 14.5% of patients. All patients had a histological diagnosis of NSCLC (77% non-squamous and 23% squamous) and were Epidermal Growth Factor Receptor (EGFR) non-mutated and Anaplastic Lymphoma Kinase gene (ALK) non-translocated. At baseline IO liver metastases were present in 35 (17.5%) of patients. More than one third of patients (35%) received IO in first line, while the remaining patients received anti-PD-(L)-1 therapy in further-lines. Overall, 40 (20%) patients were H, 65(32.5%) were I, 54 (27%) were L, according to MSC risk level. On the other hand, 41 (20.5%) patients were E and thus not analysable.

164 patients were enrolled in this study and patients were divided in two major groups. 73 belonged to the R group (CR, PR, or SD), and 91 to the NR one (PD, HPD, or NE).

Predicting Responder and Non-Responder patients

Table 2 presents the results of the feature selection procedure. The best model turns out to be the Logistic Regression which included 5 features: ECOG performance status, IO-line of therapy, the neutrophil-to-lymphocyte ratio (NLR), the MSC test and PD-L1. The parameter vector learned by LR were $w = \text{ECOG}:0.692; \text{IO LINE}: 0.71; \text{NLR}: 1.058; \text{MSC}: 0.566; \text{PD-L1}: -0.471$. This showed how an increase of

one of the first four features is negatively correlated patients response, and, conversely, the increase in the PD-L1 value correlate positively with response.

For each model, the confusion matrix is presented in Figure 3, to show their performances in terms of true/false positives/negatives.

Logistic Regression as the best model achieves ACC=0.756, F1=0.722, and AUC=0.83. PD-L1 alone has ACC=0.655 (whose performances are illustrated by the red circle in Figure 4). We also evaluated the accuracy of the LR models excluding PD-L1, MSC, and both PD-L1 and MSC from the models, i.e., considering only clinical features. Moreover, we excluded the ECOG, being the only physician-dependent feature. The results of these models are shown in Table 3, and the ROC curves are provided in Figure 1 Supplementary. Removing PD-L1, the accuracy of the corresponding model decreases to ACC=0.726 confirming the high importance of this feature, as reported in the literature. Removing the MSC from the feature decreased the accuracy to ACC=0.750, suggesting that the predictive power of this index is less impactful than PD-L1. Removing both from the data we achieve ACC=0.707. Finally, removing the ECOG decreases the accuracy of the LR model to ACC=0.726, therefore the importance of the physician clinical evaluation is comparable to PD-L1 in the prediction. These findings are confirmed by the values of the F1 score and the average AUC (Table 3). The ROC curve obtained by the leave one out methods is presented in Figure 4.

Survival analysis according to PFS and OS

Since good results have been obtained in classifying patients as responders and non-responders, it is also possible to estimate mOS and mPFS of these patients using KM curves as shown in the Figures 5(a), and 5(b). At data cut-off (Nov 2020), mOS was 10.1 months for all patients. Median PFS for R and NR group were 11.4 vs 1.8 months (HR 0.095, 95%CI 0.062 – 0.114, $p < 0.0001$) and the median OS: 38.5 vs 3.8 months (HR 0.123, 95% CI 0.079 -0.193, $p < 0.0001$).

Predicting long-survival patients (³24-months OS)

To predict if a patient is a long survival (³24-months OS) another ML binary classification analysis was performed.

Notice that since we are solving a different classification model, we need to reconsider the use of the above-mentioned methods from scratch. In Table 4 is reported all the procedures for feature selection. Even in this case the LR method resulted to be the most promising according to the AIC criterion. It achieves an ACC=0.855, F1=0.908, and AUC=0.87. The features included in the model were: ECOG, Histology, NLR and IO line.

The ROC curves computed using leave one out approach are provided in Figure 2 Supplementary.

Discussion

The use of AI is experiencing a great interest in the medical field and, in particular, in oncology. In the recent literature, there exists a wide range of publication regarding the use of AI applied to NSCLC, especially focusing on real word data, genomics, circulomics, radiomics. In our study, we aimed to find an algorithm to predict response and efficacy to IO using real word data (i.e., clinical, tumour, and treatment data) and translational ones (i.e., the result of the MSC test). Combining the current medical literature, clinical experience of the physicians, and ML tools, we developed an algorithm including 5 important features discriminating with a good accuracy (ACC = 0.756, F1 = 0.722, and AUC = 0.83) between R and NR patients. The model achieved significantly better result comparing to PD-L1 prediction value alone, which is the only currently used biomarker by physician in clinical practice to select NSCLC patients to IO that have an accuracy ACC = 0.655 on the analysed dataset. To understand if the algorithm maintains its accuracy using only real word data, we decided to exclude the PD-L1 from the model features. In this case the accuracy of the model decreased, suggesting that, even if the PD-L1 alone it is not enough to provide an effective response prediction, it remains an essential feature for IO prediction to be used in clinical practice. We did the same with the MSC, since this test is an expensive and time-consuming exam, and, therefore, its introduction in clinical practice needs to be justified. When we leave out the MSC from the model, the model accuracy reduces even if less than the case of PD-L1 exclusion, again suggesting that the MSC has a role in our model. We also tested the model removing the patient's ECOG, which it is a physician-dependent value and the results demonstrated a significant impact, analogous to PD-L1. Since the model was able to discriminate between R and NR group, we were also able to indirectly predict PFS and OS of these patients.

With a binary classification approach, we provided a method to identify and predict those patients with long OS (≥ 24 -months OS). Even in this case the use of ML techniques showed a significant improvement over the use of PD-L1 (ACC = 0.855, F1 = 0.908, and AUC = 0.87 vs. ACC = 0.734).

Different papers have been recently published to cope with our same unmet clinical need not only in NSCLC but also in other different cancer types.

Radiomics features are frequently used to predict IO response in NSCLC patients. In the study by He et al. (24) with a dual propose, radiomic were applied to build a TMB signature. CT images were used to discriminate patients with High-TMB and Low-TMB in 327 patients. The model was then applied to the IO of 123 patients' dataset to evaluate the risk stratification. The TMB radiomic signature reached an AUC of 0.74. (5) The prediction was slightly lower compared to our study, probably to indicate that the clinical features and patients' presentation has as high relevance as tumour features and it is important to consider them in the model.

Khorrami et al. (25) compared changes ("delta") in the radiomic texture of CT-scans patterns (139 patients) and associate them with tumor-infiltrating lymphocyte (TIL) density on the diagnostic biopsies in 36 patients. A linear discriminant analysis classifier yielded an AUC of 0.88 ± 0.08 in distinguishing R from NR patients when CT scan features were combined with TIL density. However, 36 patients are

included in this coupled analysis and even if our study achieved a lower AUC, our model includes 4 real world data which are easier to be obtained compared to radiomics and TIL analysis.

Yang et al. (26) used 200 patients to develop a Deep Learning (DL) model integrating different data sources (serial radiomics CT scans, laboratory and baseline clinical data) to identify R and NR subgroups to IO in NSCLC patients. The model reported an AUC of 0.80 (95% CI: 0.74–0.86), showing a smaller expected value when compared to ours (AUC 0.82). A very interesting study called DeePaN (27), used a deep patient graph convolutional network to investigate the IO benefit in NSCLC patients. By integrating real world data (age, sex, race, histology, stage, ECOG score, smoking status and previous treatment, blood analyses) and genomics in 1937 patients, the algorithm was able to divide patients in two different subgroups: beneficial and non-beneficial patients with a mOS of 20.35 and 9.42 months respectively. Comparing to our model even our sample was smaller we also were able to predict survival and response with comparable results. The model also demonstrated the positive role of TMB and KRAS mutated in IO patients (27). The study by Tian et al (28) has a dual purpose: first predicting a PD-L1 signature (PD-L1ES) using CT images (in 939 patients) and the second to predict IO response in NSCLC patients combining PD-L1ES and clinical features (in 77 patients). PD-L1ES was able to distinguish patients with a better PFS compared to those with a lower PFS. However, results of the combined model (PD-L1ES and clinical data) were superior to both the clinical and PD-L1ES models singularly (28). Our study also confirmed the importance of PD-L1 and its adding value to clinical features.

The Development and the Validation of a 12-Gene Immune Relevant Prognostic Signature for Lung Adenocarcinoma through ML strategies has been investigated in 954 patients to predict IO. From the discovery dataset of 204 observations including microarray data of gene expression of 1811 genes a Cox Regression was used to decrease the number of features to 336. Random Forest was then used to extract the final 12 selected genes used to compute the risk score. Patients were classified into high- or low-score with an AUC of 0.854, 95%CI = 0.79–0.92). Patients with a high-risk score experienced a lower survival comparing to the low one (HR = 10.6, 95%CI = 3.21–34.95, $P < 0.001$). (29)

Independently from IO, ML and DL techniques are now used in research to predict NSCLC prognosis treated with different therapies to better address precision medicine, however these techniques are still far from their introduction in clinical practice. An interesting study used DL to implement OS prediction of NSCLC patients by integrating microarray and clinical data. A list of 15 relevant genes was built using 7 known relevant biomarker genes and other less known 8 genes. Expression data of the 15 genes and the clinical data were combined and developed an integrative deep NN predicting the 5-year survival status of NSCLC patients with high accuracy (AUC: 0.8163, accuracy: 75.44%), these data are consistent and comparable with our results (30). Another study developed an algorithm to predict NSCLC survival time in 1000 patients treated with different type of therapies. Thirteen features were included in the algorithm, e.g., number of primaries, tumour size, age and stage. Random forest was the best model to predict short period survival term (< 6 months) (31).

Finally, IO biomarker prediction, as we mentioned above, is an unmet clinical need also for other cancer types. In fact, as in NSCLC different efforts have been made to find predictive biomarkers of IO response using ML or DL methodology in other cancers. An interesting report on melanoma patients integrates histologic data and clinical data to predict IO response. The algorithm consists in a segmentation classifier that takes as input the whole slide image of the patient (haematoxylin and eosin tissue). These results were then combined through a multivariable logistic regression with clinical characteristics such as age, gender, histologic subtypes, etc. The classifier accurately stratified patients into high versus low risk for disease progression with an AUC = 0.80 (32).

Gene expression data were used to separate patients in Durable Clinical Benefit (DCB) and Non-Durable Clinical Benefit (NDCB) in gastric metastatic cancer considering a Training Dataset of 25 (DCB) + 45 (NDCB) and a Validation Cohort of 9 (DCB) + 15 (NDCB), obtaining an accuracy of ACC = 0.857 on validation cohort (33).

Lastly, in another work regarding IO prediction in bladder cancer CT-scans were used to develop a ML model according to RECIST methodology and the ROI were processed to extract radiomic features. Considering a dataset of 43 subjects the model reaches an Accuracy of ACC = 0.861 (34).

Our study has different limitations: firstly, the limited sample size. Secondly, we did not use radiomic features in our study and no genomic data are included except the unique molecular data requested as for standard of care.

There are many studies that are trying to extract more information from imaging (radiomics) and genomic data. Radiomics is a very important frontier but still in an early phase and more time will need to include it in clinical practice. The same for genomics. The approach used in this paper includes routine information from imaging (e.g., RECIST) and also real world genetic data were used, those already investigated as per standard of care, which both added to the clinical can allow to better extract predictive multifactorial information. This collection can be cheaper and easier to be collected.

Conclusion

In conclusion, the results suggest that the data integration provided by AI techniques is a good tool to improve prediction for NSCLC patients treated with IO. More specifically, the model shows that higher ECOG, NLR value, IO-line, and MSC test level correlate negatively with the response to IO therapy, and, conversely, higher PD-L1 correlates positively with the response. It also confirms that PD-L1 and MSC are relevant biomarkers to improve the accuracy of the model. Moreover, considering the difference in survival among R and non-R groups, these results suggest that the model can also be used to indirectly predict survival (PFS and OS).

Finally, a second binary model was able to identify long survival patients with a high accuracy.

Abbreviations

Non-Small Cell Lung Cancer (NSCLC), Immunotherapy (IO), Programmed Death-Ligand 1 (PD-L1), Tumour Mutational Burden (TMB), tumour microenvironment (TME), microRNA (miRNA), Artificial Intelligence (AI), Machine Learning (ML), plasma microRNA signature classifier (MSC), quantitative reverse transcription PCR (RT-qPCR, low (L), intermediate (I), high (H), haemolysed (E), intravenously (IV), weeks (w); Total Body Computed Tomography (TB-CT), Response Evaluation Criteria in Solid Tumours (RECIST1.1), Complete Response (CR), Partial Response (PR), Stable Disease (SD), Progressive Disease (PD), Hyper Progression Disease (HPD), responder (R), non-responder (NR), median progression-free survival (mPFS), median overall survival (mOS), Hazard Ratio (HR), Akaike Information Criterion (AIC), *Neutrophil to Lymphocyte Ratio (NLR)*, *Lactate DeHydrogenase (LDH)*, the Accuracy (ACC), Area Under the Curve ROC (AUC), tumour-infiltrating lymphocyte (TIL).

Declarations

Ethics approval and consent to participate

This study was approved from the local ethical committee and all included patients signed informed consent prior plasma and data collection accomplished in accordance with the Declaration of Helsinki, Good Clinical Practice, and local ethical guideline.

Consent for publication

Not applicable.

Availability of data and material

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

- **MCG** declares personal financial interests with the following organizations: AstraZeneca, MSD International GmbH, BMS, Boehringer Ingelheim Italia S.p.A, Celgene, Eli Lilly, Ignyta, Incyte, Inivata, MedImmune, Novartis, Pfizer, Roche, Takeda;
- **FdB** declares: Consultant Advisory Board for Ignyta, BMS, Daiichi Sankyo, Pfizer, Octimet Oncology, Incyte, Teofarma, Pierre Fabre, Roche, EMD Serono, Sanofi, NMS Nerviano Medical Science, Pharm Research Associated (U.K) Ltd; as a Speaker BMS, Roche, MSD, Ignyta, Bayer, ACCMED, Dephaforum S.r.l., Nadirex, Merck, Biotechspert Ltd, PriME Oncology, Pfizer, Servier, Celgene, Tesaro, Loxo Oncology Inc., Sanofi, Healthcare Research & Pharmacoepidemiology, as P.I for Novartis, Roche, BMS, Celgene, Incyte, NMS, Merck KGAA, Kymab, Pfizer, Tesaro, MSD
- **AP** declares personal fees from Roche, AstraZeneca and BMS outside the submitted work.

- **CP** declares personal fees from BMS and MSD, outside the submitted work.
- **GLR** declares personal fees from BMS, MSD and Astra Zeneca outside the submitted work.
- **DS** declares personal fees from AstraZeneca, Boehringer Ingelheim and BMS, outside the submitted work.
- The other authors report no conflict of interest.

Funding

No funding was received.

Authors' contributions

- **AP**: conceptualization, investigation, methodology, project administration, supervision, writing – original draft, writing – review & editing
- **MB**: data curation, formal analysis, methodology, resources
- **AR**: data curation, formal analysis, methodology, resources, software, writing – original draft, writing – review & editing
- **RF**: data curation, formal analysis, methodology, resources, software, writing – review & editing
- **CP**: data curation, methodology, investigation, writing – review & editing
- **GLR**: data curation, methodology, investigation, writing – review & editing
- **GG**: data curation, methodology, investigation, writing – review & editing
- **ADT**: data curation, methodology, investigation, writing – review & editing
- **MB**: data curation, methodology, investigation, writing – review & editing
- **MO**: data curation, methodology, investigation, writing – review & editing
- **SM**: data curation, investigation, writing – review & editing
- **TB**: data curation, investigation, writing – review & editing
- **AB**: writing – review & editing
- **GM**: writing – review & editing
- **EZ**: data curation, investigation, writing – review & editing
- **RG**: data curation, investigation
- **EG**: data curation
- **MG**: data curation
- **GS**: data curation
- **FdB**: supervision, writing – review & editing
- **MCG**: conceptualization, investigation, methodology, supervision, writing – review & editing

- **MR:** conceptualization, data curation, formal analysis, methodology, resources, software, supervision, validation, visualization, writing – review & editing
- **ALGP:** conceptualization, data curation, formal analysis, methodology, resources, software, supervision, validation, visualization, writing – review & editing
- **FT:** conceptualization, data curation, formal analysis, methodology, resources, software, supervision, validation, visualization, writing – review & editing

Acknowledgements

We acknowledge Anna Maria Leone, our nurse for assisting us with the patients included in this study.

References

1. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68, 394–424 (2018).
2. Planchard, D. et al. Correction to: ‘Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up’. *Ann Oncol* 30, 863–870 (2019).
3. Gettinger S, Horn L, Jackman D, Spigel D, Antonia S, Hellmann M, et al. Five-Year Follow-Up of Nivolumab in Previously Treated Advanced Non-Small-Cell Lung Cancer: Results From the CA209-003 Study. *J Clin Oncol*. 2018 Jun 10;36(17):1675-1684. doi: 10.1200/JCO.2017.77.0412.
4. Reck, M. et al. Updated Analysis of KEYNOTE-024: Pembrolizumab Versus Platinum-Based Chemotherapy for Advanced Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score of 50% or Greater. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 37, 537–546 (2019).
5. Gadgeel S, Rodríguez-Abreu D, Speranza G, Esteban E, Felip E, Dómine M, et al. Updated Analysis From KEYNOTE-189: Pembrolizumab or Placebo Plus Pemetrexed and Platinum for Previously Untreated Metastatic Nonsquamous Non-Small-Cell Lung Cancer. *J Clin Oncol*. 2020 May 10;38(14):1505-1517. doi: 10.1200/JCO.19.03136.
6. Paz-Ares L, Vicente D, Tafreshi A, Robinson A, Soto Parra H, Mazières J, et al. A Randomized, Placebo-Controlled Trial of Pembrolizumab Plus Chemotherapy in Patients With Metastatic Squamous NSCLC: Protocol-Specified Final Analysis of KEYNOTE-407. *J Thorac Oncol*. 2020 Oct;15(10):1657-1669. doi: 10.1016/j
7. Pai-Scherf L, Blumenthal GM, Li H, Subramaniam S, Mishra-Kalyani PS, He K, Zhao H, et al. FDA Approval Summary: Pembrolizumab for Treatment of Metastatic Non-Small Cell Lung Cancer: First-Line Therapy and Beyond. *Oncologist*. 2017 Nov;22(11):1392-1399. doi: 10.1634/theoncologist.
8. Reck M, Schenker M, Lee KH, Provencio M, Nishio M, Lesniewski-Kmak K, et al. Nivolumab plus ipilimumab versus chemotherapy as first-line treatment in advanced non-small-cell lung cancer with high tumour mutational burden: patient-reported outcomes results from the randomised, open-label, phase III CheckMate 227 trial. *Eur J Cancer*. 2019 Jul;116:137-147. doi: 10.1016/j.ejca.2019.05.008.

9. Petitprez, F., Meylan, M., de Reyniès, A., Sautès-Fridman, C. & Fridman, W. H. The Tumor Microenvironment in the Response to Immune Checkpoint Blockade Therapies. *Front. Immunol.* 11, (2020).
10. Fortunato, O. et al. Circulating mir-320a promotes immunosuppressive macrophages M2 phenotype associated with lung cancer risk. *Int J Cancer* 144, 2746–2761 (2019).
11. Prelaj, A. et al. Predictive biomarkers of response for immune checkpoint inhibitors in non-small-cell lung cancer. *Eur J Cancer* 106, 144–159 (2019).
12. Sivan, A. et al. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* 350, 1084–9 (2015).
13. Yoon, H. J. et al. Deciphering the tumor microenvironment through radiomics in non-small cell lung cancer: Correlation with immune profiles. *PloS One* 15, e0231227 (2020).
14. Prelaj, A. et al. EPSILoN: A Prognostic Score Using Clinical and Blood Biomarkers in Advanced Non-Small-cell Lung Cancer Treated With Immunotherapy. *Clin. Lung Cancer* 0, (2020).
15. Sorich, M. J., Rowland, A., Karapetis, C. S. & Hopkins, A. M. Evaluation of the Lung Immune Prognostic Index for Prediction of Survival and Response in Patients Treated With Atezolizumab for NSCLC: Pooled Analysis of Clinical Trials. *J. Thorac. Oncol.* 1440–1446 (2019).
16. Triberti, S., Durosini, I. & Pravettoni, G. A “Third Wheel” Effect in Health Decision Making Involving Artificial Entities: A Psychological Perspective. *Front. Public Health* 8, (2020).
17. Boeri M, Verri C, Conte D, et al. MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. *Proc Natl Acad Sci U S A* 2011;108:3713-8.
18. (Sozzi G, Boeri M, Rossi M, et al. Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study. *J Clin Oncol* 2014;32:768-73.
19. Boeri M, Milione M, Proto C, Signorelli D, Lo Russo G, Carlotta Galeone C. et al. Circulating miRNAs and PD-L1 Tumor Expression Are Associated with Survival in Advanced NSCLC Patients Treated with Immunotherapy: a Prospective Study. *Clin Cancer Res.* 2019 Apr 1;25(7):2166-2173
20. Prelaj A, Proto C, Lo Russo G, Signorelli D, Ferrara R, Mensah M, et al. Integrating clinical and biological prognostic biomarkers in patients with advanced NSCLC treated with immunotherapy: the DEMo score system. *Transl Lung Cancer Res.* 2020 Jun;9(3):617-628. doi: 10.21037/tlcr-20-231.
21. Mensah M, Borzi C, Verri C, et al. MicroRNA Based Liquid Biopsy: The Experience of the Plasma miRNA Signature Classifier (MSC) for Lung Cancer Screening. *J Vis Exp* 2017;(128):56326.
22. Ferrara R, Mezquita L, Texier M, Lahmar J, Audigier-Valette C, Tessonier L, et al. Hyperprogressive Disease in Patients With Advanced Non-Small Cell Lung Cancer Treated With PD-1/PD-L1 Inhibitors or With Single-Agent Chemotherapy. *JAMA Oncol.* 2018 Nov 1;4(11):1543-1552. doi: 10.1001/jamaoncol.
23. Lo Russo G, Moro M, Sommariva M, Cancila V, Boeri M, Centonze G, et al. Antibody-Fc/FcR Interaction on Macrophages as a Mechanism for Hyperprogressive Disease in Non-small Cell Lung

- Cancer Subsequent to PD-1/PD-L1 Blockade. *Clin Cancer Res.* 2019 Feb 1;25(3):989-999. doi: 10.1158/1078-0432.CCR-18-1390.
24. B. He et al., "Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker," *J. Immunother. Cancer*, vol. 8, no. 2, pp. 1–10, 2020, doi: 10.1136/jitc-2020-000550.
 25. M. Khorrami, P. Prasanna, A. Gupta, P. Patil, P. D. Velu, R. Thawani, et al. Changes in CT Radiomic Features Associated with Lymphocyte Distribution Predict Overall Survival and Response to Immunotherapy in Non-Small Cell Lung Cancer. *Cancer Immunol Res* , 8(1):108{119, 01 2020.April 26,
 26. Y. Yang, J. Yang, L. Shen, J. Chen, L. Xia, B. Ni, L. Ge, et al. A multi-omics-based serial deep learning approach to predict clinical outcomes of single-agent anti-PD-1/PD-L1 immunotherapy in advanced stage non-small-cell lung cancer. *Am J Transl Res* , 13(2):743{756, 2021.
 27. Chao Fang Ch, Dong Xu D, Jing Su J, Jonathan R Dry JR & Bolan Linghu B. DeePaN: deep patient graph convolutional network integrating clinico-genomic evidence to stratify lung cancers for immunotherapy. *npj Digital Medicine* volume 4, Article number: 14 (2021)
 28. Tian P, He B, Mu W, Kunqin Liu K, Liu L, Zeng H et al. Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images. *Theranostics.* 2021; 11(5): 2098–2107. doi: 10.7150/thno.4802
 29. L. Xue, G. Bi, C. Zhan, Y. Zhang, Y. Yuan, and H. Fan, "Development and Validation of a 12-Gene Immune Relevant Prognostic Signature for Lung Adenocarcinoma Through Machine Learning Strategies," *Front. Oncol.*, vol. 10, no. May, pp. 1–14, 2020, doi: 10.3389/fonc.2020.00835.
 30. Y. H. Lai, W. N. Chen, T. C. Hsu, C. Lin, Y. Tsao, and S. Wu, "Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020, doi: 10.1038/s41598-020-61588-w
 31. J. A. Bartholomai and H. B. Frieboes. Lung Cancer Survival Prediction via Machine Learning Regression Classification, and Statistical Techniques. *Proc IEEE Int Symp Signal Proc Inf Tech* , 2018:632{637, Dec 2018.
 32. Johannet P, Coudray N, Donnelly DM, Jour G, Illa-Bochaca I, Xia Y et al. Using Machine Learning Algorithms to Predict Immunotherapy Response in Patients with Advanced Melanoma *Clin Cancer Res.* 2021 Jan 1;27(1):131-140. doi: 10.1158/1078-0432.CCR-20-2415.
 33. Z. Lu et al., "Prediction of immune checkpoint inhibition with immune oncology-related gene expression in gastrointestinal cancer using a machine learning classifier," *J. Immunother. Cancer*, vol. 8, no. 2, pp. 1–5, 2020, doi: 10.1136/jitc-2020-000631.
 34. F. Rundo, C. Spampinato, G. L. Banna, and S. Conoci, "Advanced deep learning embedded motion radiomics pipeline for predicting anti-PD-1/PD-L1 immunotherapy response in the treatment of bladder cancer: Preliminary results," *Electron.*, vol. 8, no. 10, 2019, doi: 10.3390/electronics8101134.

Tables

Table 1. Features selected based on literature review and clinician experience, and keeping only one for each pair of variables showing linear correlation > 0.8 .

Features Classes	Features
Clinical Features	Age, Sex, Smoker/Non-smoker, Pack per year, ECOG
Laboratory exams	NLR, NLR4, LDH
Tumor features	PD-L1, Histology
Radiological	Metastatic sites (Liver, Brain, Bone)
Treatment features	IO line
Omic features	MSC

Table 2. Features selected for the different models and corresponding performances.

	Selected features	AIC	ACC	F1	AUC
LR	ECOG, IOline, NLR, MSC, PD-L1	132.5	0.756	0.722	0.83 (0.76-0.88)
FFNN	NLR, IOline, MSC, LDH, ECOG, PackYear	137.2	0.732	0.686	0.80 (0.73-0.86)
K-NN	NLR, IOline, ECOG, MSC, NLR4	137.4	0.726	0.667	0.81 (0.74-0.87)
SVM	ECOG, IOline, NLR, MSC, PD-L1	134.5	0.738	0.703	0.83 (0.75-0.88)
RF	NLR, IOline, ECOG, Age, MSC	135.5	0.701	0.657	0.82 (0.73-0.87)

Table 3. Performances of the LR method when some of the features are removed from the initial pool of available ones.

Initial feature set	Selected Features	ACC	F1	AUC
All	ECOG, IOline, NLR, MSC, PD-L1	0.756	0.722	0.83 (0.76-0.88)
No PD-L1	ECOG, IOline, NLR, MSC	0.726	0.696	0.82 (0.75-0.88)
NO MSC	ECOG, IOline, NLR, PD-L1, Age	0.750	0.709	0.81 (0.74-0.87)
NO PD-L1 and MSC	ECOG, IOline, NLR, Age	0.707	0.662	0.80 (0.73-0.86)
NO ECOG	IOline, NLR, MSC, PD-L1	0.726	0.690	0.80 (0.73-0.87)

Table 4. Features selected for the different models and corresponding performances for the task of predicting the long-survival patients.

	Selected features	AIC	ACC	F1	AUC
LR	ECOG, Histology, NLR, IOLine	58.1	0.855	0.917	0.89 (0.80-0.94)
FFNN	Histology, NLR, PD-L1, NLR4	61.4	0.839	0.908	0.87 (0.78-0.92)
K-NN	NLR, PD-L1, Histology	60.6	0.847	0.916	0.88 (0.80-0.93)
SVM	Age, Histology, MSC, ECOG, PD-L1, NLR	63.2	0.847	0.913	0.90 (0.83-0.94)
RF	NLR, PD-L1	63.8	0.847	0.917	0.83 (0.74-0.89)

Supplementary

Supplementary Figures 1 and 2 are not available with this version

Figures

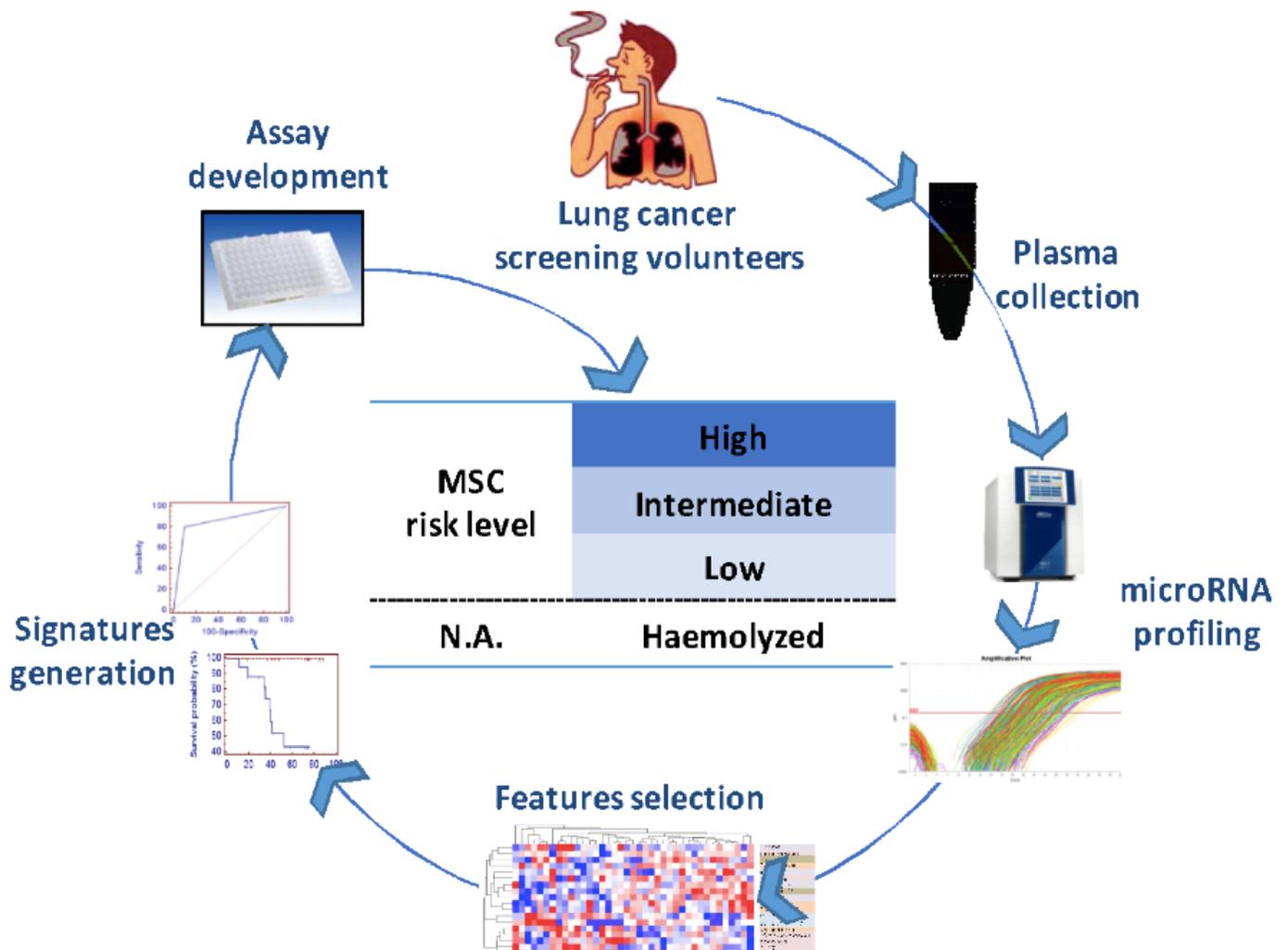


Figure 1

Development of the plasma MSC test using 24 miRNA and 4 risk groups. MSC: microRNA signature classifier; N.A.: not analyzable

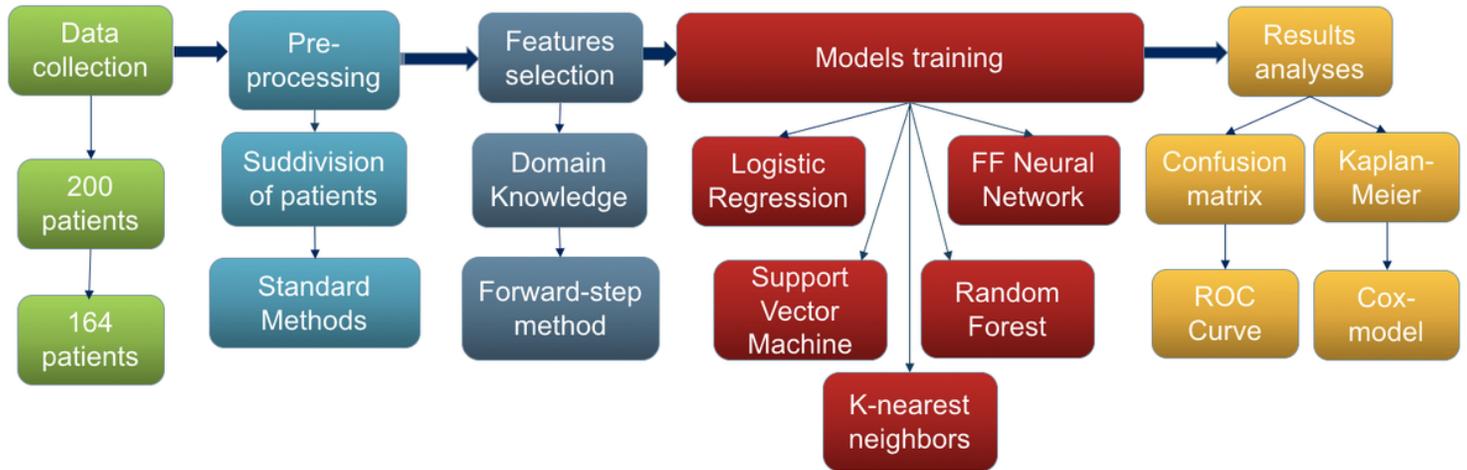


Figure 2

Process and methods used in this study.

True Class	NR	72	25
	R	19	48
		NR	R
		Predicted Class	

(a) NN

True Class	NR	72	21
	R	19	52
		NR	R
		Predicted Class	

(b) LR

True Class	NR	74	28
	R	17	45
		NR	R
		Predicted Class	

(c) K-NN

True Class	NR	70	22
	R	21	51
		NR	R
		Predicted Class	

(d) SVM

True Class	NR	68	26
	R	23	47
		NR	R
		Predicted Class	

(e) RF

Figure 3

Confusion matrix for the analysed ML models for Responders (R) and Non-Responders (NR).

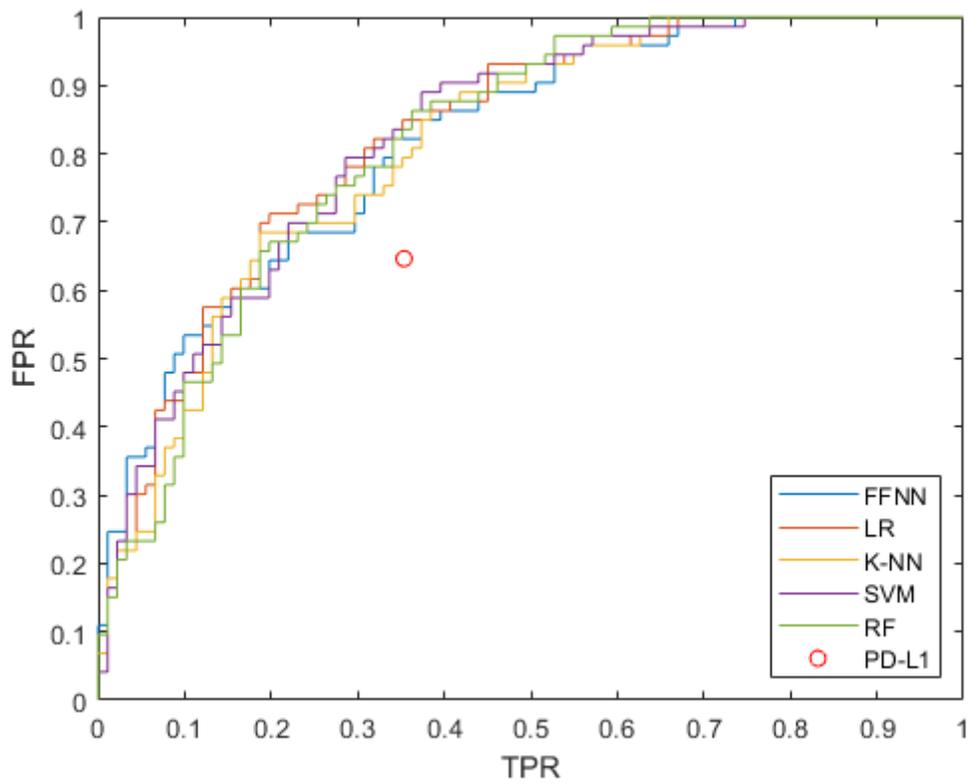
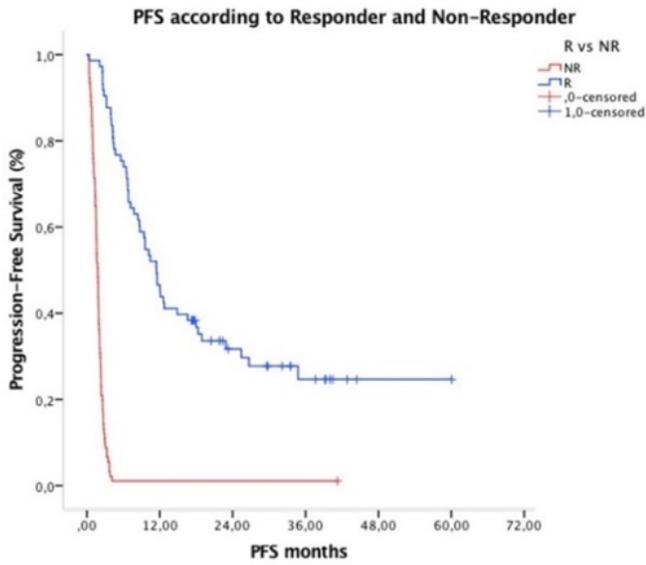
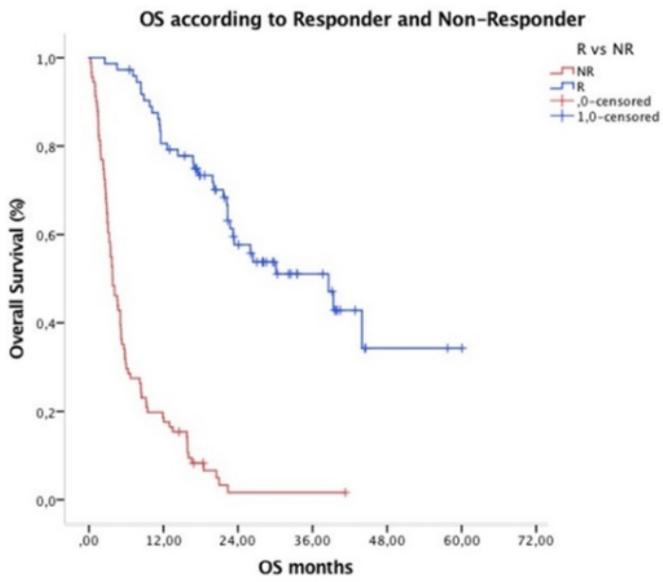


Figure 4

ROC curves (True Positive Rates (TPR) vs. False Positive Rate (FPR)) for the analysed ML models. The performance of PD-L1 is represented as a red circle. As suggested by the AUC confidence intervals, there is no method that outperforms the others significantly.



(a) PFS



(b) OS

Figure 5

Kaplan-Meier according to PFS (a) and OS (b) curves divided R (red curves) and NR (blue curves) groups.