

Machine learning for prediction of immunotherapy efficacy in non-small cell lung cancer from simple clinical and biological data

Sébastien Benzekry^{1, *}, Mathieu Grangeon^{2, *}, Mélanie Karlsen¹, Maria Alexa¹, Isabella Bicalho-Frazeto¹, Solène Chaleat², Pascale Tomasini^{2,3}, Dominique Barbolosi¹, Fabrice Barlesi^{3,4}, Laurent Greillier^{1,2}

1. COMPUTational pharmacology and clinical Oncology Department, Inria Sophia Antipolis – Méditerranée, Cancer Research Center of Marseille, Inserm UMR1068, CNRS UMR7258, Aix Marseille University UM105, Marseille, France

2. Multidisciplinary Oncology and Therapeutic Innovations Department, Assistance Publique - Hôpitaux de Marseille, Aix Marseille University, Marseille, France.

3. Aix Marseille University, CNRS, INSERM, CRCM, Marseille, France

4. Gustave Roussy Cancer Campus, International Center of Thoracic Cancers, Université Paris-Saclay, Villejuif, France.

* = these authors contributed equally

The authors declare no conflict of interest.

Social media handles:

Authors : @SBenzekry, @barlesi, @LGreillier

Institutions : @aphm_actu, @inria_sophia, @crcm_marseille

Highlights:

- Machine learning applied to a large set of NSCLC patients could predict efficacy of immunotherapy with a 69% accuracy using simple routine data
- Hemoglobin levels and performance status were the strongest predictors and significantly associated with DCR, PFS and OS
- Neutrophils-to-lymphocyte ratio was also associated with outcome
- Benchmark of 8 machine learning models

ABSTRACT

Background

Immune checkpoint inhibitors (ICIs) are now a therapeutic standard in advanced non-small cell lung cancer (NSCLC), but strong predictive markers for ICIs efficacy are still lacking. We evaluated machine learning models built on simple clinical and biological data to individually predict response to ICIs.

Methods

Patients with metastatic NSCLC who received ICI in second line or later were included. We collected clinical and hematological data and studied the association of this data with disease control rate (DCR), progression free survival (PFS) and overall survival (OS). Multiple machine learning (ML) algorithms were assessed for their ability to predict response.

Results

Overall, 298 patients were enrolled. The overall response rate and DCR were 15.3 % and 53%, respectively. Median PFS and OS were 3.3 and 11.4 months, respectively. In multivariable analysis, DCR was significantly associated with performance status (PS) and hemoglobin level (OR 0.58, $p < 0.0001$; OR 1.8, $p < 0.001$). These variables were also associated with PFS and OS and ranked top in random forest-based feature importance. Neutrophils-to-lymphocytes ratio was also associated with DCR, PFS and OS. The best ML algorithm was a random forest. It could predict DCR with satisfactory efficacy based on these three variables. Ten-fold cross-validated performances were: accuracy 0.68 ± 0.04 , sensitivity 0.58 ± 0.08 ; specificity 0.78 ± 0.06 ; positive predictive value 0.70 ± 0.08 ; negative predictive value 0.68 ± 0.06 ; AUC 0.74 ± 0.03 .

Conclusion

Combination of simple clinical and biological data could accurately predict disease control rate at the individual level.

Keywords: Blood counts, lung cancer, response, survival, prediction, machine learning

INTRODUCTION

Immune checkpoint inhibitors (ICIs) are now a therapeutic standard in several advanced cancers, particularly in stage IV non-small cell lung cancer (NSCLC) without genetic alteration ^{1,2}. The development of ICIs is leading to treat an increasing number of patients with these expensive drugs. Even if the overall response rate is higher with ICIs than chemotherapy, it is equal to about 20% for ICIs in monotherapy ². Consequently, there are still 4 patients out of 5 with no response to single agent ICI. Thus, identification of predictive markers for ICIs efficacy is an important unmet medical need.

Biologically, ICIs mechanism of action rely on the immune system and tumor micro-environment. Tumor-infiltrating lymphocytes (TILs) are known to have different effects on survival³. Blood counts may be a surrogate marker of these TILs and reflect inflammation and adaptive immune response in lung cancer ⁴. In this respect, analysis of blood counts before the start of ICIs showed some interesting correlation with response. In a meta analysis⁵, particularly in melanomas treated with Ipilimumab, a higher lymphocyte count and relative lymphocyte count predicted better overall survival (OS), as for a higher eosinophil count and lower neutrophil count. Neutrophil to lymphocyte ratio (NLR) at baseline has also been investigated, and its decrease was associated with better OS, progression-free survival (PFS) and response ⁶. The derived NLR (dNLR = absolute neutrophils count / [white blood count – absolute neutrophils count]) is another ratio which has already been an alternative to NLR in melanomas ^{7,8} and metastatic colorectal cancer ⁹. Furthermore, for lung cancer, in a Chinese meta-analysis published in 2016 ¹⁰, high levels of platelet to lymphocytes ratio (PLR) at baseline were associated with poor OS and PFS, but in all types of treatment. Specifically for NSCLC treated by ICIs, a study showed that a score combining dNLR greater than 3 with elevated LDH was correlated with worse outcome for ICIs ¹¹. Furthermore, an Italian study in NSCLC patients treated by ICIs ¹² showed that low NLR and low PLR at baseline were associated with development of immune related adverse events (IRAEs), and low NLR was associated with better outcomes (OS, PFS). However, a comprehensive analysis of all classical blood markers for prediction of efficacy in a large number of patients is still lacking ¹³.

In the era of precision medicine, machine learning (ML) has recently developed as an alternative to classical statistical analysis¹⁴. The main difference is that statistical analysis focuses on inference and association between variable(s) and outcome(s), while ML puts emphasis on predictive performances only¹⁵. In oncology, ML has demonstrated great success for prediction from large-dimensional ‘big’ data, such as genomics or imaging¹⁴. Nevertheless, such data science methods also have relevance to establish predictive models from smaller sets of variables¹⁶. In addition, successes have mostly been limited to diagnosis and prognosis but seldom for predictive applications in a clinical oncology context (i.e., for therapeutic decision).

We hypothesized that ML could be useful to accurately predict the efficacy of ICIs in NSCLC patients. The present study aimed to develop a ML model for the selection of patients which could benefit from treatment with ICIs, from simple clinical and biological data.

MATERIALS AND METHODS

Patients

In this observational monocentric retrospective study, we analyzed data from all patients older than 18 years of age who were diagnosed with advanced NSCLC and who received at least 1 cycle of ICI (anti-PD-L1, anti PD-1 or anti-CTLA-4), alone or in combination (anti-PD(L)1 and anti-CTLA4) following a first cycle of non-ICI systemic therapy from April 2, 2013 to February 14, 2017. Patients were treated according to available guidelines.

The study protocol and retrospective data collection were approved by the Institutional Review Board of the French Society of Respiratory Diseases (Société de Pneumologie de Langue Française – SPLF), under reference number: CEPRO 2019-007 and patients have signed an informed consent.

Data

Data were retrieved from electronic patient records. Clinical and epidemiological data (age, gender, tobacco status, asbestos exposure, performance status, body mass index), disease characteristics (histology, mutation status, TNM stage), treatment data (type, treatment line, toxicity), biological data (last blood count before first infusion of ICI) and outcome data (response and survival) were collected. From the pre-treatment blood counts, we calculated the PLR, NLR, and derived NLR (dNLR) = absolute neutrophils count / (white blood count – absolute neutrophils count). Performance status was dichotomized into < 2 or ≥ 2 .

Tumor response was assessed once every 2 months through computed tomography scans, according to the Response Evaluation Criteria in Solid Tumors, version 1.1.¹⁷ Definition of response here is the best response observed. Overall response rate (ORR) included patients with complete or partial responses. Disease control (DCR) included patients with complete, partial responses or stable disease. Overall survival was defined as the time from start of immunotherapy to death from any cause, censored at the date of last follow-up. Progression-free survival was defined as the time from start of immunotherapy to documented disease progression or death from any cause, censored at the date of last follow-up period.

Statistical Analysis

In exploratory data analysis, two-tailed Student's t-tests were used for continuous variables and chi-squared tests for categorical variables. Association of clinical and biological data response was assessed using univariable and multivariable logistic regression, as implemented in the *glm* function of the *R* software (version 3.6)¹⁸. Survival analyses of OS and PFS were performed using univariable and multivariable proportional hazard Cox's regression models^{19,20}. Continuous variables were centered and scaled before these analyses.

Machine Learning

Feature selection was performed using feature importance given by a random forest classification algorithm applied on the entire data set (*randomForest* *R* package, no tuning, 1,000 trees). Once sorted by mean decrease accuracy, incremental logistic regression models were built with increasing number of variables. The selected optimal set of features was the maximal one before observing a decrease in 10-fold cross-validated accuracy. Predictive machine learning (ML) models were further built and assessed using repeated nested k-fold cross-validation with 5 repeats of 3 outer loops (to assess generalizability) containing each 5 repeats of 3 inner loops (to tune the models). Thus, 15 train and test sets were built to test the predictive performances. The models were implemented under the *tidymodels* framework in *R* version 4.0.4²¹. They included: logistic regression (*glm*), random forest (*ranger*, 1000 trees), single layer neural network (2600 maximum number of weights), naïve Bayes, k-nearest neighbors (*kkn*, rectangular kernel function) and support vector machines (with linear, polynomial or radial basis kernel function). For each outer fold, models were tuned and trained on the train set and predictions were assessed on the test set. The decision tree was built using the *rpart* engine and tuned for hyper-parameters of tree depth, minimum number of data points required for further splitting (*min_n*) and complexity parameter (*cost_complexity*), using a grid search and a 10-fold cross-validation. The tree was then trained on the entire set.

RESULTS

Patients and Disease Characteristics

Overall, 298 patients treated with ICIs for stage IV or relapsed NSCLC were retrieved from our database and analyzed. Patient and disease characteristics are summarized in Table 1. Regarding the treatments with ICIs, 89% (n = 266) of patients received an anti-PD-1 antibody, with 96.7% (n = 286) patients being pretreated with chemotherapy prior to ICIs. The report of blood counts values at start of treatment with ICIs are reported in Table S1.

Statistical Analysis

Response

The DCR was 53.4% and ORR was 15.4%. Exploratory data analysis was conducted for association of the considered variables with outcome (Figure 1). Significant associations were found for NLR ($p < 0.001$), derived NLR ($p < 0.001$), hemoglobin ($p < 0.0001$), leukocytes ($p < 0.01$) and neutrophils ($p < 0.001$). These results were confirmed by logistic regression analysis, with additional significance of (Table 2). However, in multivariable analysis, only hemoglobin and PS remained significant. They also remained significant in multivariable analysis controlling for possible additional confounding factors: sex, immunotherapy type and smoking status.

Survival analysis

Progression-free and overall survival are reported in Figure S1. The median PFS was 3.27 months (95%CI: 2.63 - 4.07) and median OS was 14.5 months (95%CI: 8.8 – 15.5). Proportional hazard Cox regression confirmed association of hemoglobin and performance status with response. They were significantly correlated with PFS and OS, in univariable and multivariable analysis (Tables S2 and S3). They also remained significant in multivariable analysis controlling for possible additional confounding factors: sex, immunotherapy type and smoking status.

Machine Learning

The ML analysis was conducted for prediction of DCR (classification task) and comprised two steps. First, feature selection and then ML classification. The first step was conducted using random forest-based mean decrease in accuracy (Figure 2A) followed by selection of an optimal number of predictors (Figure 2B). The former

revealed hemoglobin level as the strongest predictor of DCR. The second strongest predictor was performance status, followed by NLR. Adding further predictors resulted in a decrease in the cross-validated accuracy of logistic regression models (Figure 2B). Thus, these three variables were selected for further inclusion in ML models.

Multiple machine learning models using this set of variables were then assessed for their predictive abilities of DCR (Figure 3 and Table 3). First, learning curves –which evaluate the models predictive abilities with increasing number of patients – demonstrated that convergence to the optimal predictive power had been reached, for each model (Figure S2). Receiver-operator curves were similarly discriminant across the algorithms (Figure 3A), apart from k-nearest neighbors (knn) and the polynomial support vector machine (SVM) with poorer performances. Aside knn, mean areas under the curve ranged 0.72 – 0.74 (Table 3). Similarly, precision (sensitivity) – recall (positive predictive value) curves were comparable (Figure 3B). Best accuracy was 68%, achieved by the random forest model. Sensitivity was generally low (max 0.58, random forest) while specificity was high (0.73 – 0.94, Table 3). Best positive (naive Bayes, 0.72) and negative (random forest, 0.68) predictive values suggested good predictive power (Table 3). Altogether, the ML algorithms performances suggested the random forest algorithm as the most adequate, achieving highest score in the largest number of them (accuracy, area under the ROC curve, sensitivity and negative predictive value, Table 3) and exhibiting the smallest inter-score variability (Figure 3C).

A random forest algorithm being hard to interpret, we also trained a decision tree algorithm (Figure 3D). This confirmed performance status (<2 versus ≥ 2) and hemoglobin level (optimal threshold 13 g/dL) as the most important predictive variable. Then NLR, consistently with our random forest-based importance analysis. This tree could be useful for clinical decision at bedside. For instance, patients with patient status < 2 and hemoglobin ≥ 13 g/dL are predicted to have an 83% chance of disease control.

DISCUSSION

The selection of patients who will benefit from ICIs therapies is crucial in the era of precision medicine, in order to develop new strategies for those patients who are not likely to respond from current strategies. Clinical examination and blood counts are easily acquired, but their predictive power in combination remains to be determined. While classical statistical methods are appropriate and have been employed for determination of association with outcome ^{11,22}, ML techniques are more adapted for prediction tasks. Therefore, we decided to analyze our data with the help of such methods.

Although well developed in several areas of science and industry, especially for dealing with ‘big data’, the use of ML for clinical oncology has remained limited to date ^{14,23}. In particular, very few studies have investigated ML for prediction of response to immune-checkpoint blockade, and none has focused on the predictive value of blood counts ^{16,24–26}. In addition, the main limitation of such studies is the small sample size, despite being a critical determinant to ensure the robustness and generalizability of the results ²⁷. For instance, Wiesweg et al. had 55 patients in the training set and 36 in the test set ²⁵.

In comparison, we analyzed the data from 298 patients, allowing to have higher statistical power. Training and test sets were thus composed of about 200 and 100 patients in training and tests, respectively. This large number of patients ensured that the models had enough information to learn and predict (Figure S2). Our results even show that 200 patients are sufficient to reach the optimal accuracy, for models with six variables. Importantly, our data was collected from clinical practice (i.e., real-world data), which implies larger heterogeneity but is also more reflective of the reality at bedside ²². The random forest algorithm emerged as the algorithm with best trade-off over all the metrics considered, resulting in a 68% mean accuracy and 0.74 mean area under the ROC curve, on test sets. Nevertheless, logistic regression, single layer neural network and naïve Bayes models performed almost equally well, suggesting that similar predictive power can be achieved with a standard, linear statistical learning algorithm such as logistic regression. This is in line with a review that suggested no benefit of ML over logistic regression for clinical prediction models ²⁸.

Associations between blood count and efficacy were consistent with previous studies, in particularly correlations with hemoglobin, NLR, dNLR and PLR, were the same than in previous studies ^{5,6,10,11,29}. This consistency comforts the interest of pre-treatment blood counts for prediction of ICIs efficacy. Neutrophils and leukocytes levels were also associated and predictive of response, a finding from our study that has not been reported elsewhere, to our knowledge. We also demonstrated that an ECOG performance status ≥ 2 was significantly associated with response, a result that remained to be fully established ^{30,31}. High PS and a high NLR were correlated with worse outcomes, consistently with the physio-pathological immune hypothesis of peripheral lymphocytes stimulation with ICIs, which could lead to redirect TILs to destruct tumor cells through an activation cascade.

The present study focused on the predictive power of blood biomarkers but the model could include additional variables, even if their predictive values are not perfect individually. PD-L1 expression ³², because of its discordant results, its heterogeneity among histological components ³³ and its poor accuracy when assessed in peripheral blood ³⁴, is not an ideal predictive biomarker but would probably increase the performance of our model. Some characteristics based on genomic alterations could also be added. Tumor mutational burden (TMB), defined as the total number of nonsynonymous mutations in the tumor exome ³⁵ and evaluated by next-generation-sequencing (NGS), seems to be a potential biomarker ³⁶. Even if its measure has some limitation such as high cost, large input DNA amount needed and time of analysis, lung cancer is within tumors with the highest TMB ³⁷, which correlates with ICI efficacy in NSCLC³⁸. Importantly, our results compare favorably with the latter study, where predictive power of (durable) response from PD-L1 expression (AUC = 0.646) or TMB (AUC = 0.601) was inferior. K-RAS mutation is a frequent genetic alteration with contrasting implications for ICIs efficacy. It was initially not linked with poor response to ICIs ³⁹, but a better response when mutated was shown in 2019 ⁴⁰, whereas the co-occurring genomic alteration of K-RAS and STK11/LKB1 was associated with a primary resistance to PD1 axis blockade⁴¹. Other recent work using metabolomics data obtained from blood sampling demonstrated impressive predictive power of response, although with small sample sizes ^{42,44}. The integration of such markers in our model might improve its efficacy. Based on the combination of quantitative imaging and machine learning, radiomics study have also started to emerge for prediction of ICIs

efficacy from radiological images^{24,42}. Such studies should nevertheless be carefully evaluated for their added predictive power in comparison to simpler biomarkers as the ones used here⁴³.

The first limitation of our study is the retrospective design and the limited number of centers involved (2 centers). Another limitation is the lack of tumor PD-L1 assessment. This testing was not systematically performed from 2013 to 2017 as nivolumab in second-line or more did not require the PD-L1 status to be prescribed to NSCLC patients. This status could have helped our algorithm using a quantitative variable, freeing ourselves from a qualitative threshold as currently used : 1% for the refund of Durvalumab in France for the maintenance after a radio-chemotherapy or pembrolizumab in second-line therapy⁴⁴ and 50% for the use of Pembrolizumab in first-line therapy². Nevertheless, the PD-L1 determination methods were not the same as nowadays, involving either a different or biased analysis or a new determination of PD-L1 status.

Certain confounding factors such as bacterial and viral infection, demographic variables such as race, recent chemotherapy, or the use of corticosteroids before treatment could have modified the blood counts analyzed in this study. These data were not collected and not analyzed in our study because we wanted a simple tool to help clinical practice, which could benefit to the majority of patients. Furthermore, blood counts were not necessarily sampled the day of first ICIs infusion, but sometimes one or some days before. This period of time could also have modified the values. On the other hand, our study reflects clinical practice in real-life.

Eventually, the majority of studies to date, including ours, have focused on static pre-treatment predictive markers. However, dynamic markers that integrate on-treatment data start to emerge, with promising predictive value⁴⁵. In such context, the use of statistical Bayesian modeling is particularly appropriate and holds great promise⁴⁶. For ICI, a mathematical model-derived kinetic parameters of tumor kinetics regrowth during relapse has been shown to be the best predictor of overall survival in multivariable analysis including baseline clinical markers⁴⁷. More mechanistic mathematical models for tumor-immune-ICI-radiotherapy dynamics have also been proposed⁴⁸. In light of our results, such models should be applied to include dynamics of blood counts.

CONCLUSION

Blood counts prior to ICIs (elevation of hemoglobin, decrease of NLR, leukocytes or neutrophils) and clinical status (good PS) were significantly associated with better DCR in multivariable analysis. The practical application of these associations using machine learning algorithms was able to predict individual response to treatment. This could be improved further by increasing the number of variables in the model and should be further validated in an independent cohort.

ACKNOWLEDGEMENTS

The authors would like to thank Pauline Fleury for valuable input on the work. This work is part of the QUANTIC project funded by ITMO Cancer AVIESAN and French Institut National du Cancer (grant #19CM148-00)

References

1. 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.
2. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med* 2016;375(19):1823–1833.
3. Gooden MJM, Bock GH de, Leffers N, Daemen T, Nijman HW. The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis. *Br J Cancer* 2011;105(1):93–103.
4. Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet* 2001;357(9255):539–545.
5. Hopkins AM, Rowland A, Kichenadasse G, et al. Predicting response and toxicity to immune checkpoint inhibitors using routinely available blood and clinical markers. *Br J Cancer* 2017;117(7):913–920.
6. Diem S, Schmid S, Krapf M, et al. Neutrophil-to-Lymphocyte ratio (NLR) and Platelet-to-Lymphocyte ratio (PLR) as prognostic markers in patients with non-small cell lung cancer (NSCLC) treated with nivolumab. *Lung Cancer Amst Neth* 2017;111:176–181.
7. Proctor MJ, McMillan DC, Morrison DS, Fletcher CD, Horgan PG, Clarke SJ. A derived neutrophil to lymphocyte ratio predicts survival in patients with cancer. *Br J Cancer* 2012;107(4):695–699.
8. Ferrucci PF, Ascierto PA, Pigozzo J, et al. Baseline neutrophils and derived neutrophil-to-lymphocyte ratio: prognostic relevance in metastatic melanoma patients receiving ipilimumab. *Ann Oncol* 2016;27(4):732–738.
9. Diakos CI, Tu D, Gebski V, et al. Is the derived neutrophil to lymphocyte ratio (dNLR) an independent prognostic marker in patients with metastatic colorectal cancer (mCRC)? Analysis of the CO.17 and CO.20 studies. *Ann Oncol* [Internet] 2016 [cited 2019 Jul 23];27(suppl_6). Available from: https://academic.oup.com/annonc/article/27/suppl_6/588P/2799240
10. Gu X, Sun S, Gao X-S, et al. Prognostic value of platelet to lymphocyte ratio in non-small cell lung cancer: evidence from 3,430 patients. *Sci Rep* [Internet] 2016;6. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4812293/>
11. Mezquita L, Auclin E, Ferrara R, et al. Association of the Lung Immune Prognostic Index with Immune Checkpoint Inhibitor Outcomes in Patients with Advanced Non-Small Cell Lung Cancer. *JAMA Oncol* 2018;4(3):351–357.
12. Pavan A, Calvetti L, Dal Maso A, et al. Peripheral Blood Markers Identify Risk of Immune-Related Toxicity in Advanced Non-Small Cell Lung Cancer Treated with Immune-Checkpoint Inhibitors. *The Oncologist* 2019;
13. Barlesi F, Greillier L, Monville F, et al. LBA53 Precision immuno-oncology for advanced non-small cell lung cancer (NSCLC) patients (pts) treated with PD1/L1 immune checkpoint inhibitors (ICIs): A first analysis of the PIONeeR study. *Ann Oncol* 2020;31:S1183.
14. Benzekry S. Artificial intelligence and mechanistic modeling for clinical decision making in oncology. *Clin Pharmacol Ther* 2020;108(3):471–486.
15. Breiman L. Statistical modeling: the two cultures. *Stat Sci Rev J Inst Math Stat* 2001;16(3):199–231.
16. Ahn B-C, So J-W, Synn C-B, et al. Clinical decision support algorithm based on machine learning to assess the clinical response to anti-programmed death-1 therapy in patients with non-small-cell lung cancer. *Eur J Cancer* 2021;153:179–189.

17. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45(2):228–247.
18. R: The R Project for Statistical Computing [Internet]. [cited 2019 Oct 2]; Available from: <https://www.r-project.org/>
19. Cox DR. Regression Models and Life-Tables. *J R Stat Soc Ser B Methodol* 1972;34(2):187–202.
20. Therneau TM. A Package for Survival Analysis in R [Internet]. 2021. Available from: <https://CRAN.R-project.org/package=survival>
21. Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. [Internet]. 2020. Available from: <https://www.tidymodels.org>
22. Becker T, Weberpals J, Jegg AM, et al. An enhanced prognostic score for overall survival of patients with cancer derived from a large real-world cohort. *Ann Oncol* 2020;31(11):1561–1568.
23. Nicolò C, Périer C, Prague M, et al. Machine Learning and Mechanistic Modeling for Prediction of Metastatic Relapse in Early-Stage Breast Cancer. *JCO Clin Cancer Inform* 2020;4:259–274.
24. Trebeschi S, Drago SG, Birkbak NJ, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol* 2019;30(6):998–1004.
25. Wiesweg M, Mairinger F, Reis H, et al. Machine learning-based predictors for immune checkpoint inhibitor therapy of non-small-cell lung cancer. *Ann Oncol* 2019;30(4):655–657.
26. Wiesweg M, Mairinger F, Reis H, et al. Machine learning reveals a PD-L1–independent prediction of response to immunotherapy of non-small cell lung cancer by gene expression context. *Eur J Cancer* 2020;140:76–85.
27. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441.
28. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;110:12–22.
29. Zhang Z, Zhang F, Yuan F, et al. Pretreatment hemoglobin level as a predictor to evaluate the efficacy of immune checkpoint inhibitors in patients with advanced non-small cell lung cancer. *Ther Adv Med Oncol* 2020;12:1758835920970049.
30. Dall’Olio FG, Maggio I, Massucci M, Mollica V, Fragomeno B, Ardizzoni A. ECOG performance status ≥ 2 as a prognostic factor in patients with advanced non small cell lung cancer treated with immune checkpoint inhibitors-A systematic review and meta-analysis of real world data. *Lung Cancer Amst Neth* 2020;145:95–104.
31. Krishnan M, Kasinath P, High R, Yu F, Teply BA. Impact of Performance Status on Response and Survival Among Patients Receiving Checkpoint Inhibitors for Advanced Solid Tumors. *JCO Oncol Pract* 2021;OP.20.01055.
32. Melosky B, Chu Q, Juergens RA, et al. Breaking the biomarker code: PD-L1 expression and checkpoint inhibition in advanced NSCLC. *Cancer Treat Rev* 2018;65:65–77.
33. Liu Y, Dong Z, Jiang T, et al. Heterogeneity of PD-L1 Expression Among the Different Histological Components and Metastatic Lymph Nodes in Patients With Resected Lung Adenosquamous Carcinoma. *Clin Lung Cancer* 2018;19(4):e421–e430.
34. Boffa DJ, Graf RP, Salazar MC, et al. Cellular Expression of PD-L1 in the Peripheral Blood of Lung Cancer Patients is Associated with Worse Survival. *Cancer Epidemiol Prev Biomark* 2017;26(7):1139–1145.
35. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500(7463):415–421.

36. Berland L, Heeke S, Humbert O, et al. Current views on tumor mutational burden in patients with nonsmall cell lung cancer treated by immune checkpoint inhibitors. *J Thorac Dis* 2018;11(1):S71–S80.
37. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer genes. *Nature* 2013;499(7457):214–218.
38. Rizvi H, Sanchez-Vega F, La K, et al. Molecular Determinants of Response to Anti–Programmed Cell Death (PD)-1 and Anti–Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non–Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. *J Clin Oncol* [Internet] 2018 [cited 2021 Aug 10]; Available from: <http://ascopubs.org/doi/pdf/10.1200/JCO.2017.75.3384>
39. Jeanson A, Tomasini P, Souquet-Bressand M, et al. Efficacy of Immune Checkpoint Inhibitors in KRAS-Mutant Non-Small Cell Lung Cancer (NSCLC). *J Thorac Oncol* 2019;14(6):1095–1101.
40. Mazieres J, Drilon A, Lusque A, et al. Immune checkpoint inhibitors for patients with advanced lung cancer and oncogenic driver alterations: results from the IMMUNOTARGET registry. *Ann Oncol* 2019;30(8):1321–1328.
41. Skoulidis F, Goldberg ME, Greenawalt DM, et al. STK11/LKB1 Mutations and PD-1 Inhibitor Resistance in KRAS-Mutant Lung Adenocarcinoma. *Cancer Discov* 2018;8(7):822–835.
42. Sun R, Limkin EJ, Vakalopoulou M, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol* 2018;19(9):1180–1191.
43. Welch ML, McIntosh C, Haibe-Kains B, et al. Vulnerabilities of radiomic signature development: The need for safeguards. *Radiother Oncol J Eur Soc Ther Radiol Oncol* 2019;130:2–9.
44. Herbst RS, Baas P, Kim D-W, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet Lond Engl* 2016;387(10027):1540–1550.
45. Mezquita L, Preeshagul I, Auclin E, et al. Predicting immunotherapy outcomes under therapy in patients with advanced NSCLC using dNLR and its early dynamics. *Eur J Cancer Oxf Engl 1990* 2021;151:211–220.
46. Kurtz DM, Esfahani MS, Scherer F, et al. Dynamic Risk Profiling Using Serial Tumor Biomarkers for Personalized Outcome Prediction. *Cell* 2019;178(3):699-713.e19.
47. Claret L, Jin JY, Ferté C, et al. A Model of Overall Survival Predicts Treatment Outcomes with Atezolizumab versus Chemotherapy in Non-Small Cell Lung Cancer Based on Early Tumor Kinetics. *Clin Cancer Res* 2018;24(14):3292–3298.
48. Serre R, Benzekry S, Padovani L, et al. Mathematical Modeling of Cancer Immunotherapy and Its Synergy with Radiotherapy. *Cancer Res* 2016;76(17):4931–4940.

Figures and Tables

Table 1: Patients and disease characteristics

Figure 1: Exploratory data analysis

Table 2: Logistic regression analysis for disease control

Figure 2: Variable selection

Figure 3: Machine learning algorithms predictive performances

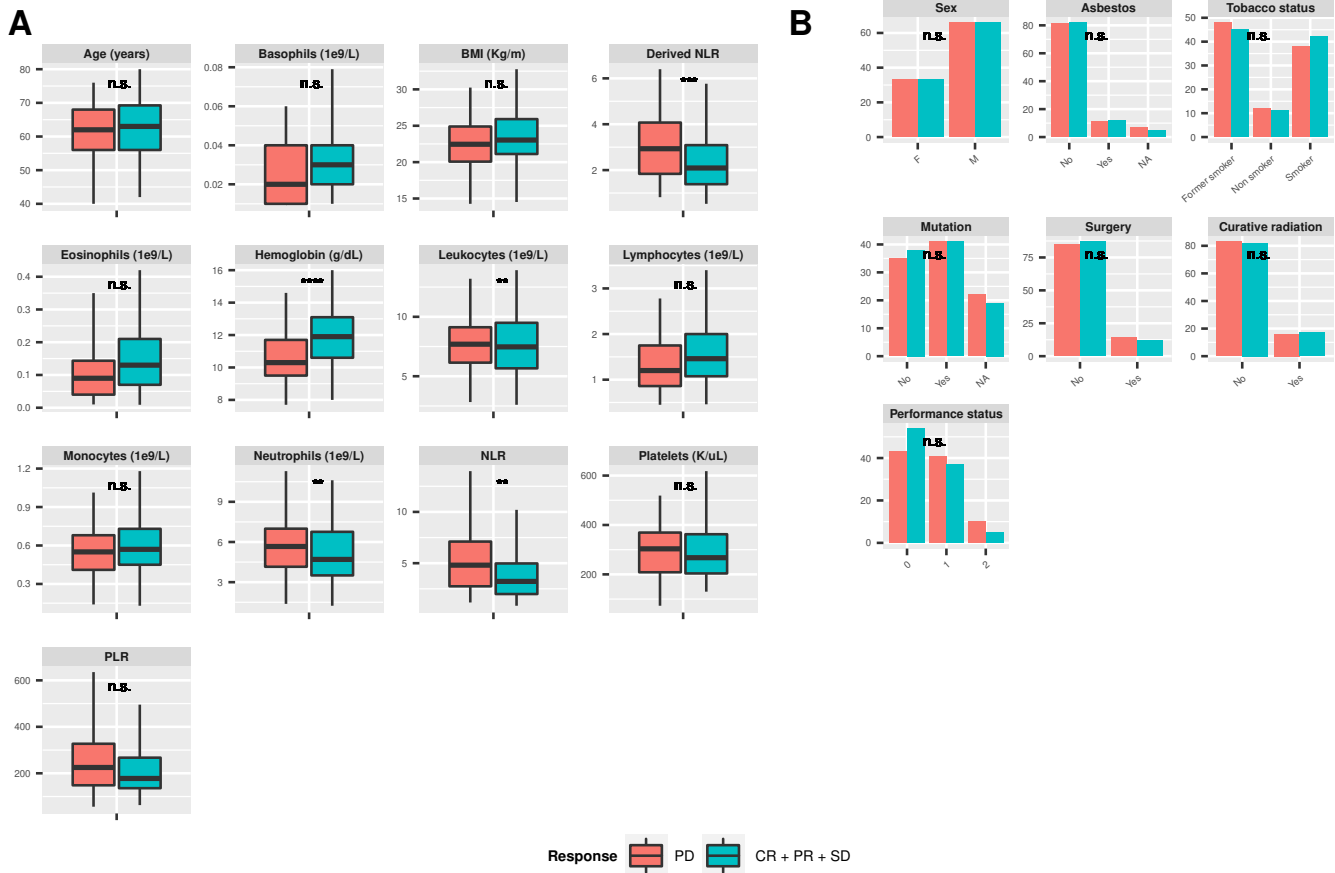
Table 3: Summary of machine learning algorithms predictive performances

Table 1: Patients and disease characteristics

Variable	N = 298 ¹
Age	62 (55, 69)
Sex	
Female	99 (33%)
Male	199 (67%)
Tobacco status	
Former Smoker	140 (47%)
Non smoker	36 (12%)
Smoker	122 (41%)
Brain metastases	72 (29%)
Performance status	
≥ 2	26 (8.9%)
0-1	265 (91%)
Mutation profile	
ALK	1 (0.8%)
BRAF	9 (7.4%)
EGFR	14 (11%)
KRAS	87 (71%)
Other mutation	5 (4.1%)
ROS1	2 (1.6%)
Wild type	4 (3.3%)
Immunotherapy type	
anti-CTLA-4	3 (1.0%)
anti-PD-1	266 (89%)
anti-PD-L1	26 (8.7%)
Combination	3 (1.0%)
Previous treatment	
Chemotherapy	281 (95%)
Chemotherapy + immunotherapy	5 (1.7%)
Targeted therapy	11 (3.7%)
Response	
Complete response	2 (0.7%)
Partial response	44 (15%)
Progressive disease	131 (45%)
Stable disease	113 (39%)

¹Median (IQR); n (%)

Figure 1: Exploratory data analysis



A. Boxplots of continuous variables

B. Barplots of categorical variables

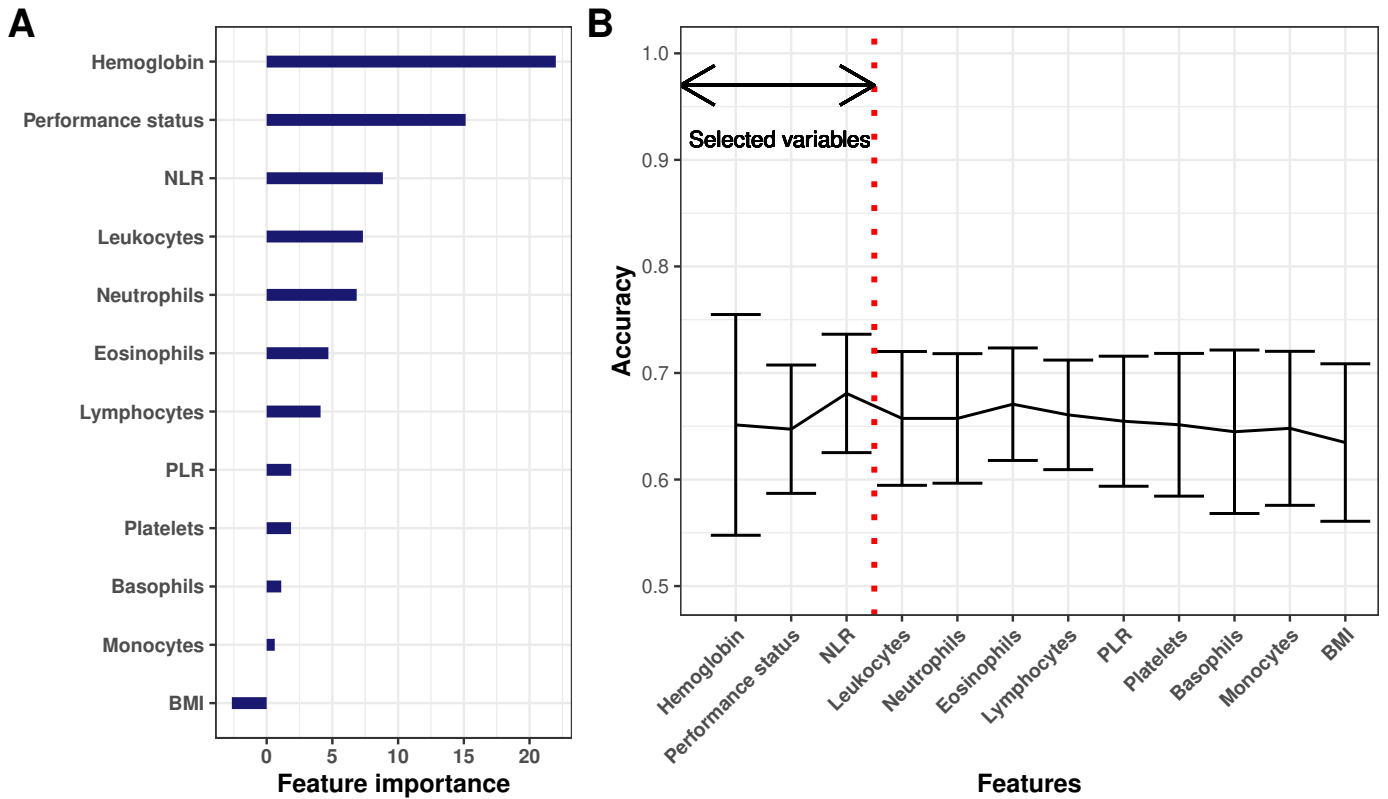
BMI = body mass index, NLR = neutrophils-to-lymphocytes ratio, PLR = platelets-to-lymphocytes ratio, CR = complete response, PR = partial response, SD = stable disease and PD = progressive disease. Stars indicate statistical significance:

* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$, n.s. = non significant.

Table 2: Logistic regression analysis for disease control

Variable	Univariable logistic regression		Multivariable logistic regression	
	Odds ratio [95% IC]	p signif	Odds ratio [95% IC]	p signif
Lymphocytes	1.1 [0.83, 1.4]	0.678	0.98 [0.15, 5.2]	0.984
NLR	0.49 [0.31, 0.73]	0.000879 ***	0.68 [0.098, 1.9]	0.651
Platelets	1 [0.82, 1.3]	0.762	1.3 [0.72, 2.4]	0.404
PLR	0.84 [0.64, 1.1]	0.156	1.1 [0.5, 2.4]	0.788
Leukocytes	0.68 [0.5, 0.89]	0.00791 **	0.6 [0.0022, 3e+02]	0.847
Hemoglobin	1.9 [1.5, 2.5]	9.26e-07 ***	1.8 [1.3, 2.4]	0.000122 ***
dNLR	0.63 [0.47, 0.83]	0.00155 **	0.8 [0.33, 2.7]	0.689
Neutrophils	0.62 [0.45, 0.83]	0.00232 **	1.5 [0.0047, 2.7e+02]	0.863
Monocytes	0.87 [0.69, 1.1]	0.226	0.86 [0.5, 1.4]	0.545
Eosinophils	1.3 [0.97, 1.9]	0.139	1.1 [0.75, 1.8]	0.582
Basophils	1.2 [0.95, 1.8]	0.177	1.2 [0.89, 1.8]	0.321
BMI	1.2 [0.95, 1.5]	0.123	1 [0.76, 1.3]	0.997
Performance status	0.5 [0.39, 0.64]	6.21e-08 ***	0.58 [0.44, 0.75]	7.79e-05 ***

Figure 2: Variable selection

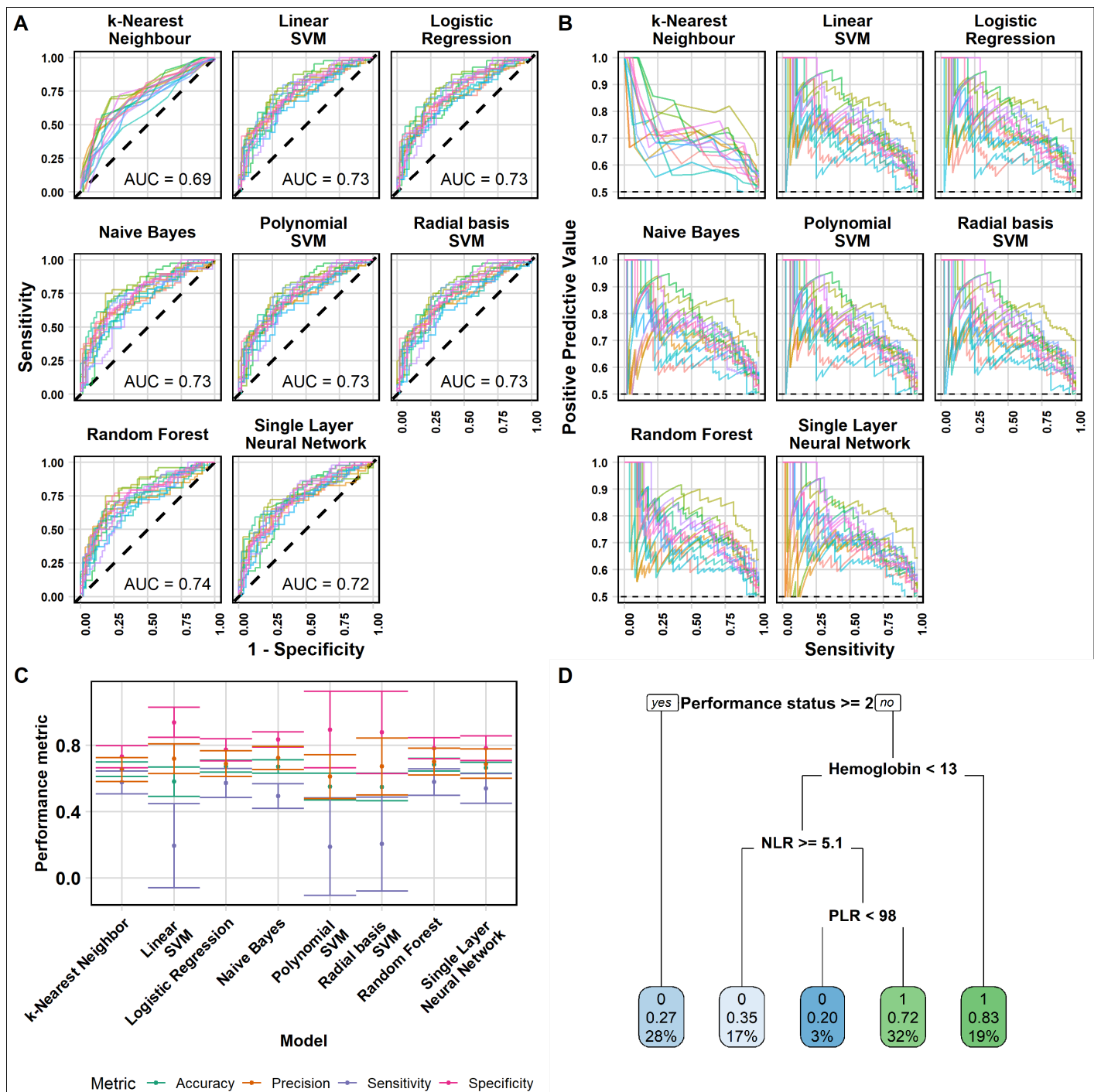


A. Feature importance based on random forest classification and mean decrease in accuracy.

B. Accuracy score of incremental logistic regression models built on an increasing number of predictors (i.e., the first one contains only hemoglobin, the second hemoglobin and NLR, etc..).

NLR = neutrophils-to-lymphocyte ratio. PLR = platelets-to-lymphocytes ratio. BMI = body mass index

Figure 3: Machine learning algorithms predictive performances



A. Receiver-operator curves for prediction on test sets from each fold of the outer cross-validation loop, for each model. AUC = area under the curve.
 B. Precision (positive-predictive value) - recall (sensitivity) curves.
 C. Main performance metrics for each algorithm.
 D. Decision tree obtained after tuning and training. Each node shows: the predicted class (0 = PD, 1 = CR+PR+SD), the predicted probability of response and the percentage of total observations in the node. PD = progressive disease, CR = complete response, PR = partial response, SD = stable disease.

Table 3: Summary of machine learning algorithms predictive performances

Model	Accuracy	ROC AUC	PPV	NPV	Sensitivity	Specificity
Random Forest	0.68 $\hat{\pm}$ 0.04	0.74 $\hat{\pm}$ 0.03	0.70 $\hat{\pm}$ 0.08	0.68 $\hat{\pm}$ 0.06	0.58 $\hat{\pm}$ 0.08	0.78 $\hat{\pm}$ 0.06
Logistic Regression	0.67 $\hat{\pm}$ 0.04	0.73 $\hat{\pm}$ 0.03	0.69 $\hat{\pm}$ 0.08	0.67 $\hat{\pm}$ 0.06	0.57 $\hat{\pm}$ 0.09	0.77 $\hat{\pm}$ 0.07
Naive Bayes	0.67 $\hat{\pm}$ 0.04	0.73 $\hat{\pm}$ 0.03	0.72 $\hat{\pm}$ 0.07	0.65 $\hat{\pm}$ 0.06	0.49 $\hat{\pm}$ 0.07	0.83 $\hat{\pm}$ 0.05
Single Layer Neural Network	0.66 $\hat{\pm}$ 0.03	0.72 $\hat{\pm}$ 0.03	0.69 $\hat{\pm}$ 0.09	0.66 $\hat{\pm}$ 0.06	0.54 $\hat{\pm}$ 0.09	0.78 $\hat{\pm}$ 0.07
k-Nearest Neighbour	0.66 $\hat{\pm}$ 0.04	0.69 $\hat{\pm}$ 0.04	0.65 $\hat{\pm}$ 0.07	0.66 $\hat{\pm}$ 0.06	0.58 $\hat{\pm}$ 0.07	0.73 $\hat{\pm}$ 0.07
Linear SVM	0.58 $\hat{\pm}$ 0.09	0.73 $\hat{\pm}$ 0.03	0.72 $\hat{\pm}$ 0.09	0.58 $\hat{\pm}$ 0.10	0.19 $\hat{\pm}$ 0.25	0.94 $\hat{\pm}$ 0.09
Polynomial SVM	0.55 $\hat{\pm}$ 0.08	0.73 $\hat{\pm}$ 0.03	0.61 $\hat{\pm}$ 0.13	0.58 $\hat{\pm}$ 0.13	0.19 $\hat{\pm}$ 0.29	0.89 $\hat{\pm}$ 0.23
Radial basis SVM	0.55 $\hat{\pm}$ 0.08	0.73 $\hat{\pm}$ 0.03	0.67 $\hat{\pm}$ 0.17	0.56 $\hat{\pm}$ 0.06	0.20 $\hat{\pm}$ 0.28	0.88 $\hat{\pm}$ 0.25

Supplementary figures and tables

Table S1: Blood counts at ICI treatment start

Figure S1: Progression-free and overall survival

Table S2: Cox regression: progression-free survival

Table S3: Cox regression: overall survival

Figure S2: Learning curves

Table S1: Blood counts at ICI treatment start

Variable	Mean	Median	Min	Max
Basophils	0.04	0.03	0.00	0.80
Derived NLR	3.15	2.37	0.41	18.58
Eosinophils	0.16	0.11	0.00	3.26
Hemoglobin	11.32	11.30	7.30	16.40
Leucocytes	8.56	7.99	2.09	49.65
Lymphocytes	1.55	1.35	0.11	19.31
Monocytes	0.63	0.58	0.07	1.96
Neutrophils	6.17	5.52	0.86	46.00
NLR	5.66	3.85	0.46	96.91
Platelets	317	290	57	2,013
PLR	273	214	3	2,309

Figure S1: Progression-free and overall survival

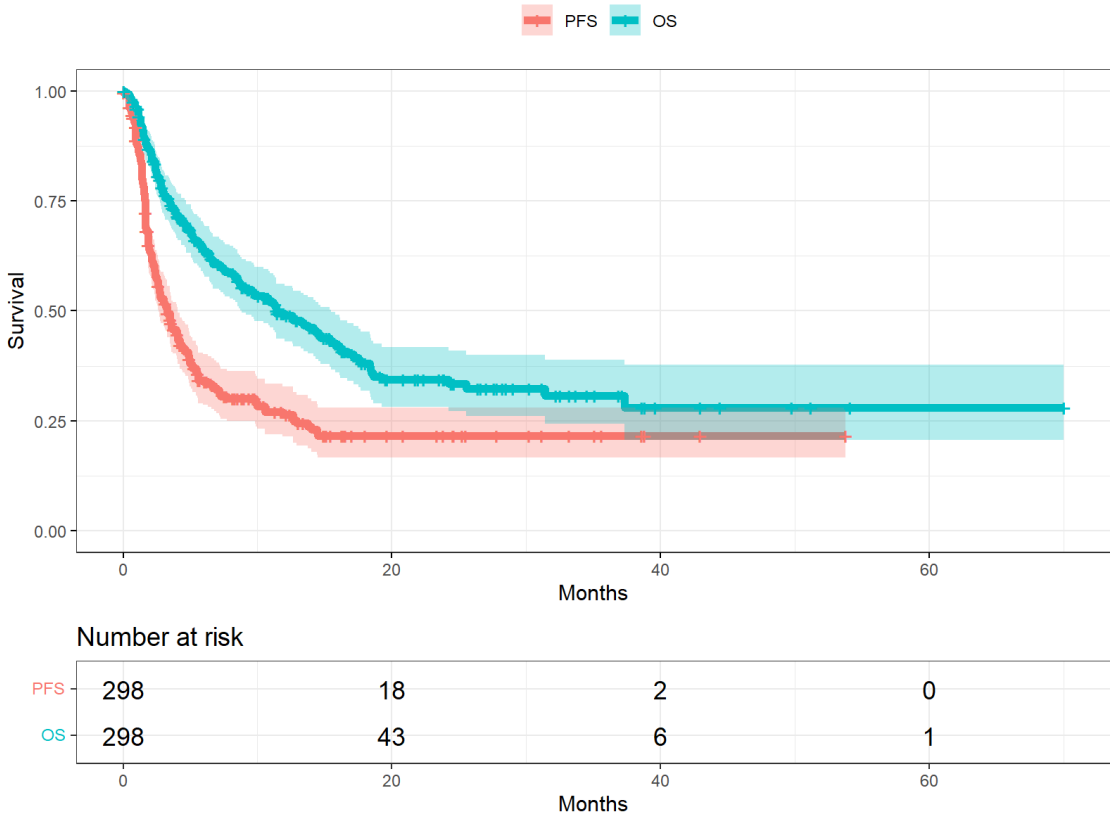


Table S2: Cox regression: progression-free survival

Variable	Univariable cox regression		Multivariable cox regression	
	Hazard ratio [95% IC]	p signif	Hazard ratio [95% IC]	p signif
Lymphocytes	0.85 [0.71, 1]	0.0912 .	1 [0.38, 2.6]	0.996
NLR	1 [1, 1]	0.00585 **	0.71 [0.33, 1.5]	0.376
Platelets	1 [1, 1]	0.37	0.81 [0.59, 1.1]	0.206
PLR	1 [1, 1]	0.881	1 [0.66, 1.5]	0.994
Leukocytes	1.1 [1, 1.1]	0.00246 **	0.69 [0.044, 11]	0.795
Hemoglobin	0.9 [0.83, 0.97]	0.00525 **	0.84 [0.72, 0.98]	0.0291 *
dNLR	1.1 [1, 1.1]	0.0017 **	1.2 [0.69, 1.9]	0.596
Neutrophils	1.1 [1, 1.1]	6.26e-05 ***	2 [0.15, 26]	0.603
Monocytes	1.1 [0.66, 1.8]	0.774	1 [0.78, 1.3]	0.885
Eosinophils	0.53 [0.23, 1.2]	0.138	0.96 [0.77, 1.2]	0.695
Basophils	0.91 [0.14, 6.1]	0.923	0.99 [0.87, 1.1]	0.917
BMI	0.96 [0.93, 0.99]	0.0133 *	0.88 [0.76, 1]	0.088 .
Performance status	2.1 [1.6, 2.9]	8.15e-07 ***	1.3 [1.1, 1.5]	0.000335 ***

Table S3: Cox regression: overall survival

Variable	Univariable cox regression		Multivariable cox regression	
	Hazard ratio [95% IC]	p signif	Hazard ratio [95% IC]	p signif
Lymphocytes	0.8 [0.64, 0.99]	0.045 *	1.1 [0.4, 3.1]	0.845
NLR	1 [1, 1.1]	4.54e-08 ***	0.87 [0.43, 1.7]	0.686
Platelets	1 [1, 1]	0.81	0.83 [0.58, 1.2]	0.305
PLR	1 [1, 1]	0.331	1 [0.65, 1.6]	0.949
Leukocytes	1.1 [1, 1.1]	6.63e-07 ***	0.99 [0.054, 18]	0.994
Hemoglobin	0.83 [0.76, 0.91]	3.58e-05 ***	0.76 [0.64, 0.92]	0.00392 **
dNLR	1.2 [1.1, 1.2]	8.93e-09 ***	1.4 [0.83, 2.2]	0.225
Neutrophils	1.1 [1.1, 1.1]	4.05e-10 ***	1.3 [0.087, 19]	0.857
Monocytes	1.4 [0.85, 2.3]	0.195	1.1 [0.82, 1.5]	0.558
Eosinophils	0.13 [0.032, 0.54]	0.00509 **	0.78 [0.55, 1.1]	0.166
Basophils	0.72 [0.071, 7.4]	0.784	0.99 [0.84, 1.2]	0.943
BMI	0.97 [0.93, 1]	0.0715 .	0.93 [0.8, 1.1]	0.405
Performance status	2.8 [2, 3.9]	1.12e-09 ***	1.4 [1.2, 1.6]	3.68e-05 ***

Figure S2: Learning curves

