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

Kyoung-Ho Pyo[a,b†], Beung-Chul Ahn[b*†], Jea-Woo So[c†], Chun-Bong Synn[a,d], Jae Hwan Kim[a], Tae Hyoung Kim[c], Chang Gon Kim[b], Sun Min Lim[b], Min Hee Hong[b], Byoung Chul Cho[b*], Hye Ryun Kim[b*]

[a]Severance Biomedical Science Institute, Yonsei University College of Medicine, Seoul, Korea
[b]Department of Internal Medicine, Division of Medical Oncology, Yonsei Cancer Center, Yonsei University College of Medicine, Seoul, Korea
[c]Theragen Etex Bio Institute, Suwon, Korea
[d]Brain Korea 21 Plus Project for Medical Science, Yonsei University, Seoul, Korea
†These authors contributed equally

Corresponding authors:

Hye Ryun Kim, MD, PhD Associate Professor

Yonsei Cancer Center, Division of Medical Oncology

Yonsei University College of Medicine

50 Yonsei-ro, Seodaemun-Gu, Seoul 120-752, Korea

Tel.: 82-2-2228-80831, Fax: 82-2-393-3562

Email: nobelg@yuhs.ac

Byoung Chul Cho, MD, PhD Professor

Yonsei Cancer Center, Division of Medical Oncology

Yonsei University College of Medicine

50 Yonsei-ro, Seodaemun-Gu, Seoul 120-752, Korea

Tel.: 82-2-2228-8126/ Fax: 82-2-393-3562

Email: cbc1971@yuhs.ac

Sources of support: This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation funded by the Ministry of Science and ICT (NRF-2019M3A9B6065231, 2017M3A9E802971, 2017M3A9E9072669 to H.R.K).

Abstract

Anti-programmed death (PD)-1 therapy confers sustainable clinical benefits for non-small cell lung cancer (NSCLC) patients, but only some patients respond to the treatment. Various clinical characteristics, including the PD-ligand 1 (PD-L1) level, are related to the anti-PD-1 response; however, none of these can independently serve as predictive biomarkers. Herein, we established a machine learning (ML)-based clinical decision support algorithm to predict the anti-PD-1 response by comprehensively combining the clinical information. We collected clinical data, including patient characteristics, mutations, and laboratory findings, from the electronic medical records of 142 NSCLC patients treated with anti-PD-1 therapy; these were analyzed for the clinical outcome as the discovery set. Nineteen clinically meaningful features were employed in supervised ML algorithms, including LightGBM, XGBoost, multilayer neural network, ridge regression, and linear discriminant analysis, to predict anti-PD-1 responses. Based on the prediction performance of each ML algorithm, the optimal ML was selected and validated in an independent validation set of PD-1 inhibitor-treated patients. PD-L1 expression, tumor burden, and neutrophil-to-lymphocyte ratio could independently predict the anti-PD-1 response in the discovery set. ML platforms based on LightGBM and XGBoost using the combination of 19 clinical features showed more significant prediction performance than on using individual clinical features. Both models showed higher significant prediction accuracy based on the cut-off value. Patients with higher cut-off values presented significantly longer median progression-free survival in the discovery and validation sets. Collectively, LightGBM and XGBoost offer a clinical decision support algorithm to predict the anti-PD-1 response in NSCLC patients.

Keywords: machine learning; clinical decision support system; lung cancer, immune checkpoint inhibitor, anti-programmed death-1, noninvasive biomarker

Introduction

Immune checkpoint inhibitors (ICIs), including programmed cell death protein 1/programmed death-ligand 1 (PD-1/PD-L1) inhibitors, have resulted in prolonged survival and were approved as first- and second-line therapies in patients with recurrent/metastatic non-small cell lung cancer (NSCLC) [1,2]. PD-L1 protein expression has been approved as a biomarker for investigating the efficacy of PD-1/PD-L1 inhibitors. PD-L1 expression is enriched in anti-PD-1/PD-L1 inhibitor therapy responders [3,4]. Nevertheless, fewer than 30% of NSCLC patients respond to anti-PD-1 inhibitors [5]. Moreover, a substantial proportion of PD-L1-positive patients show no response to therapy, while a subset of PD-L1-negative patients do show. Therefore, PD-L1 expression alone may not comprehensively reflect the complexity of the response of the tumor microenvironment to PD-1 inhibitors.

With increasing use of anti-PD-1 therapy, various clinical characteristics related to treatment response in NSCLC patients have been identified, such as neutrophil/lymphocyte ratio before immunotherapy, smoking history, performance status, sex, the presence of metastases and driver mutations, and pathology [6-14]. These factors are routinely examined in clinical practice prior to anti-PD-1 therapy and are collected in electronic medical records (EMRs). However, none of the clinical factors can accurately predict the response to anti-PD-1 therapy; thus, a model integrating these factors is needed.

Recently, machine learning (ML)-based methods have been developed to predict

disease progression and treatment response in various diseases [15,16], ML tools can identify key features from complex datasets associated with a specific purpose and interest. The ML techniques, multilayer neural network (MNN), ensemble learning, support vector machine (SVM), and penalized regression have been widely employed in recent studies for developing predictive models to facilitate effective and accurate decision-making [8,17,18]. However, ML has not yet been extensively studied or employed in ICI-treated cancer patients.

In this study, we explored and validated various predictive algorithms using an ML approach based on the clinicopathological factors of anti-PD-1-therapy–treated NSCLC patients from prospectively collected data obtained from EMRs. We aimed to establish a clinical decision support system for prescribing anti-PD-1 therapy in NSCLC patients using clinicopathological information routinely collected during clinical practice.

Materials and methods

Patients

Patients with histologically confirmed stage IV NSCLC who were treated with anti-PD-1 therapy (nivolumab 2 mg/kg every 2 weeks or pembrolizumab 200 mg fixed dose every 3 weeks) at Yonsei Cancer Center (Seoul, Korea) between March 2014 and April 2020 were included (n=192). Patients were divided into a discovery cohort—patients consecutively enrolled from March 2014 to January 2018—to explore the optimal algorithm and the validation set—patients enrolled after January 2018—to confirm the performance of the explored algorithm. Data, including patient characteristics, laboratory results, tumor size, genetic mutational status, metastatic sites, line of therapy, PD-1 inhibitor type, and response were prospectively collected from the

EMRs. The patients were treated intravenously with nivolumab at a dose of 3.0 mg/kg every 2 weeks or pembrolizumab at a fixed dose of 200 mg every 3 weeks. This study was approved by the Institutional Review Board (IRB no. 4-2016-0678) of our hospital. The need for informed consent was waived.

Determination of PD-L1 expression in tumors

PD-L1 expression was analyzed using PD-L1 IHC 22C3 pharmDx antibody (clone 22C3; Dako North America, Inc., Carpinteria, CA, USA). PD-L1 expression in tumor cells was determined based on the percentage of PD-L1-expressing cells in each section, which was estimated in increments of 5%, except for a 1% positivity value. Patients were considered positive when at least 1% of the tumor cells expressed PD-L1.

Assessment of the clinical efficacy of anti-PD1 inhibitors

The treatment response was evaluated using chest and abdomen computed tomography (CT) scanning [19]. CT scanning was performed every 3 or 4 cycles during the treatment. Besides regular follow-ups, additional imaging was conducted based on the physician's concerns. Clinical response to the PD-1 inhibitor was evaluated using the same imaging modalities and Response Evaluation Criteria in Solid Tumors (RECIST), v1.1. The primary outcome of clinical response (responder) was defined by radiographic evidence of complete response, partial response, or stable disease for at least 6 months. Lack of a clinical response (non-responder) was defined by progressive disease on serial CT scans or stable disease lasting less than 6 months. Progression-free survival (PFS) was defined as the time from the start of anti-PD-1 therapy to PD or death. Overall survival (OS) was defined as the time from the start of

anti-PD-1 therapy to death from any cause. Patients were censored on April 2020 if alive and progression-free. Patients without a known date of death were censored at the time of last follow-up. Anti-PD-1 therapy-related adverse events were reported according to the Common Terminology Criteria for Adverse Events, v4.0.

Clinical features and ML models

We extracted clinicopathological factors known to be related to anti-PD-1 therapy through a comprehensive literature review. To avoid compounding effects, features dependent on other features were removed. We modified some variables to increase explanatory power and reduce complexity **(Table S1)**. These selected features were used as variables for establishing an ML model. The best-performing ML techniques and associated hyperparameters were selected using leave-one-out cross-validation (LOOCV). We performed standardization and normalization for models that were affected by the data scale. The predictive models XGBoost, LightGBM, MNN, SVM, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), ridge regression, and lasso regression were compared and used for analyses [20-26]. The relatively small sample size (n=192) was expected to be sensitive to any noise or randomness of partitioning data; therefore, we performed LOOCV to avoid high variance and select an appropriate model.

Feature attribution analysis

A local surrogate model was used to calculate the expected value of the contribution of each feature to the prediction. The additive feature attribution method proposed by Lundberg and Lee [27] was used:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i',$$

where $z' \in \{0,1\}^M, M$ is the number of simplified features, ϕ_i is the contribution of the feature *i*, and *g* is the local explanation model minimized difference between $g(z')$ and $f(h_x(z'))$. h_x maps simplified inputs to the original input, and when x' is the simplified input of x and $x = h_x(x')$, the explanation model $g(x')$ matches the original model $f(x)$. We used Shapley values for ϕ_i for uniqueness and consistency of allocated contribution values:

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!\,|M - |z'| - 1|!}{M!} [f_x(z') - f_x(z' \backslash i)]$$

Here, the LightGBM model had the highest accuracy; therefore, the contribution value was calculated directly through the Tree Explainer algorithm, and the contribution values of each weak leaner were combined using the law of aggregation property of the SHAP value.

Results

Patient characteristics

In total, 192 patients with advanced NSCLC treated with anti-PD-1 therapy were included for analysis. Majority of patients were male (148/192, 77.1%), had adenocarcinoma (130/192, 67.7%), and were smokers (135/192, 70.3%); the median age was 64 years (range, 26–85 years). Thirty-one (17.6%) patients had driver mutations, including mutations in epidermal growth factor receptor (*EGFR*; n=25), anaplastic lymphoma kinase (*ALK*; n=2), and proto-oncogene tyrosine-protein kinase 1 (*ROS1*; n=4). Regarding lines of therapy, 15 (7.0%), 77 (40.1%), and 100 (88.9%) patients received anti-PD-1 therapy as first-line, second-line, and more than second-

line therapy, respectively. The most frequently developed metastatic site at baseline was the ipsilateral or contralateral lung (137/192, 71.4%), followed by the brain (69/192, 35.9%), bone (56/192, 29.2%), and adrenal gland (34/192, 17.7%).

The discovery (n=142) and validation cohorts (n=50) showed no significant differences in patient characteristics, including age, performance status, driver mutation status, line of therapy, and site of baseline metastasis (**Table 1**).

Table 1. Baseline characteristics of the independent discovery and validation sets.

	Discovery set N (%)	Validation set N (%)	
Age (years)			
Median	64 (26–85)	64 (38-82)	0.57
Sex			
Male	101 (71.1)	47 (94.0)	
Female	41 (27.1)	3 (6.0)	<0.05
ECOG PS score			
0	20 (14.1)	4 (8.0)	
1	87 (61.3)	39 (78.0)	
2	20 (14.1)	6 (12.0)	
3	15 (10.6)	1 (2.0)	0.109
Smoking			
Never	50 (35.2)	7 (14.0)	
Former smoker	55 (38.7)	41 (82.0)	
Current smoker	37 (26.1)	2 (4.0)	<0.05
Histology			
Adenocarcinoma	100 (70.4)	30 (60.0)	
Squamous	40 (28.2)	18 (36.0)	0.248
Others	2 (1.4)	2 (4.0)	
EGFR, ALK, and ROS1			
Wild-type (all)	114 (80.3)	43 (86.0)	
Mutant	25(17.6)	6 (12.0)	0.352
Unknown	3 (2.1)	1 (2.0)	
PD-L1			
Positive	94 (66.2)	37 (74.0)	
Negative	38 (26.8)	13 (26.0)	
Unknown	10 (7.0)	0 (0.0)	0.145
Prior treatment lines			
0	14 (9.9)	1 (2.0)	
1	51 (35.9)	26 (52.0)	
2	30 (21.1)	11 (22.0)	
≥3	47 (33.1)	12 (24.0)	0.097
Location of metastasis			
Lung ipsilateral	99 (63.4)	29 (58.0)	0.50
Lung contralateral	74 (52.1)	23 (46.0)	0.457

Brain	56 (39.4)	13 (26.0)	0.089
Bone	43 (30.3)	13 (26.0)	0.567
Adrenal gland	28 (19.7)	6 (12.0)	0.219
Liver	22 (15.6)	8 (16.0)	0.932

*Female and never smoker status had a phi coefficient of 0.839 ($p < 0.001$). § One *ALK* rearrangement and four *ROS1* rearrangements were included; ¶One *ALK* rearrangement was included.

Treatment outcomes in the discovery set and role of PD-L1 as a predictive biomarker

In the discovery set, 56 responders and 86 non-responders were classified based on prespecified definitions, and a duration of response of 6 months was observed (**Figure 1A**). The responder group contained patients with a long duration of response **(Figure 1A)**. The median PFS and OS (calculated from the date of anti-PD-1 treatment) for all patients were 8.28 and 13.54 months, respectively (**Figure 1B**). The PFS according to response was 22.14 (95% confidence interval[CI]=4.165–8.164) and 3.79 months (95%CI=0.123–0.24) in responders and non-responders, respectively. Patients with more than 50% PD-L1 expression showed a significantly prolonged PFS compared with patients with less than 50% PD-L1 (21.76 months [95%CI=1.713–3.799] versus 8.53 months [95%CI=0.263–0.583]; *P*=0.002).

Figure 1. PD-L1 expression is not a promising predictive biomarker of anti-PD-1 therapy response in NSCLC patients. (A) Swimmer's plot of anti-PD-1-treated patients with advanced NSCLC (stage 4; n=138). The left heatmap shows results of RECIST (red, partial response; gray, stable disease; blue, progressive disease), and the right heatmap shows the results of RECIST-based responses (orange, responder; black, non-responder). The colors of the lines and symbols are labeled in the symbol index (lower bottom index). **(B)** Kaplan-Meier plot for progression-free survival (PFS) in responders, non-responders, and all patients. Statistical analysis was performed

using the log-ranked method (****P < 0.0001). (C) Kaplan-Meier plot for PFS based on PD-L1 expression (≥50% and < 50%). **P < 0.01.

Overall study development scheme

To establish an optimal prediction algorithm for anti-PD-1 therapy, we collected datasets, tested multiple ML algorithms, and conducted cross-validation in the discovery set. We also used selected algorithms with an independent validation set (**Figure 2**). After collecting all available clinical features originating from EMRs, we selected the relevant clinical features by knowledge-based clinical insights and feature selection methods. The pre-feature selection process revealed 19 of the 60 features to be associated with clinical outcome **(Table S1)**. Preprocessing was used to prepare these 19 features for ML by normalization, categorization, and outlier exclusion. We selected and then tested nine types of ML models as clinical decision-supporting systems for predicting disease progression. The prediction performances of these nine ML models were compared, and two optimal prediction models were chosen. An additional 50 NSCLC patients were evaluated using the two ML algorithms to assess anti-PD-1 responses (**Figure 2**).

Figure 2. Dataset generation and application of ML methods for prediction of anti-PD-1 responses in patients with NSCLC. The dataset was generated from radiological, hematological, histological, and clinical features. The clinically important features were filtered according to ontological meanings, and asymmetric features were removed. All features were normalized and categorized, and some outliers were removed. In total, 138 patients in the discovery dataset were analyzed by ML

algorithms, including deep learning, LightGBM, random forest, SVM, ridge/lasso regressions, and logistic regression. The performance of individual ML algorithms was compared based on area under the receiver operating characteristic curve (AUC) values and cross-validation [10-fold, leave-one-out cross-validation (LOOCV)]. The validation dataset was organized (n=50), and responses to anti-PD-1 were predicted.

Clinical feature selection for ML

Heatmaps were used to visualize 10 representative features related to treatment outcomes for anti-PD-1 therapy in the discovery cohort **(Figure 3A).** Several single clinical features were found to be enriched in the responders. There were significant correlations between PD-L1 expression (cut-off 50%) and response (*P*=0.016) (**Figure 3B**). The positive predictive value (PPV, response rate above the cut-off) was 55.3% (26/47), whereas the negative predictive value (NPV, nonresponse rate below the cut-off) was 67.0% (55/82). These findings suggested that predictability could be further improved. PD-L1 expression, tumor burden and neutrophil-to-lymphocyte ratio (NLR) were significantly related to the anti-PD-1 response (**Figure 3C, D**). In two examples of discrepancy between PD-L1 expression and clinical response, one patient without PD-L1 expression (PD-L1=0%) showed a favorable response (PFS: 35.7 months, OS: 49.3 months; **Figure 3E**), whereas another patient with high PD-L1 expression (PD-L1, 100%) showed a poor response (PFS: 1.2 months, OS: 2.17 months; **Figure 3F**).

Figure 3. Clinical features and anti-PD-1 responses. (A) Anti-PD-1 responder prediction demonstrated by a heatmap (red, responder; black, non-responder). Individual prediction scores were divided by Youden index-based cut-off values. **(B)** Scores for the LightGBM model were statistically analyzed (Fisher's exact test, $p > 0.0001$). **(C)** Scores for the XGBoost model were statistically analyzed (Fisher's exact

test, $p > 0.0001$). **(E)** A representative CT image of an anti-PD-1 responder showing no PD-L1 expression in the tumor. **(F)** A representative CT image of an anti-PD-1 non-responder showing PD-L1 expression in the tumor.

ML-based prediction of anti-PD-1 responders

We hypothesized that the comprehensive integration of clinical features rather than a single clinical feature may provide important clues to predict anti-PD-1 treatment outcomes. Therefore, we attempted to develop an optimal prediction model by integrating selected clinical features using ML algorithms. We tested the performance of nine ML models, including LightGBM, XGBoost, MNN, ridge regression, LDA, Gaussian process, SVM, lasso regression, and QDA, using the discovery cohort.

Figure 4A lists all ML models derived from our discovery set in order of accuracy. The heatmap shows the estimated anti-PD-1 responses in NSCLC patients in the discovery set. Patients are indicated as responders or non-responders using black or red bars based on the specific cut-off value for each model. The Youden index was used as cut-off value. Area under the curve (AUC) values were used for all models (**Figure S1A, B**). The top two most accurate prediction models were derived from LightGBM and XGBoost algorithms, which both showed significant correlations. The AUC of the two models were 0.834 and 0.824, respectively. Both models significantly predicted responders when the cut-off was 0.452 and 0.476, respectively (**Figure 4B, C**). The PPV was 73.4% (47/64) and 73.0% (46/63) and the NPV was 88.4% (69/78) and 87.3% (69/79) in LightGBM and XGBoost models, respectively. These two models also showed significant differences between PFS, indicating that our models could be used for predicting the durability of treatment.

Both models showed higher significant prediction accuracy by cut-off value ($P < 0.001$, PPV=73.4%, NPV=88.4% in LightGBM, $P < 0.001$, PPV=73.0%, NPV=88.4% in XGBoost). The patients with a cut-off ≥ 0.452 in LightGBM and ≥ 0.476 in XGBoost presented a significantly longer median PFS than those with lower cut-off value (20.5 vs. 3.38 months, $P < 0.0001$ in LightGBM, 20.5 vs. 3.62 months $P < 0.0001$ in XGBoost) in the discovery cohort (**Figure 4D**). The cut-off level for these models could discriminate PFS in the response to anti-PD-1 therapy.

Figure 4. Machine learning-based prediction of anti-PD-1 responders. (A) Individual features analyzed and reported as SHAP values and interaction scores using the LightGBM model. Feature prediction values are shown as different colors (high, yellow; low, blue). **(B)** Hierarchical clustering of feature correlation. Individual features were analyzed (top two major clusters are indicated by green and red). **(C)** Heatmap of feature correlation. Highly correlated groups are indicated in yellow, whereas groups with less strong correlations are indicated in black. **(D)** Kaplan-Meier plot of anti-PD-1-treated patients with NSCLC using LightGBM and XGBoost model-based prediction in the discovery set. The red line indicates the predicted responder group and the black line indicates the non-responder group. The left plot is for LightGBM and the right plot is for XGBoost. The log-rank methods were used for statistical analyses between responder and non-responder groups (****$p < 0.0001$).

Feature contribution in the LightGBM model

Each of the 19 selected clinical features was respectively analyzed to determine its contribution to LightGBM. The most prominent feature was on the top, and the least prominent feature was on the bottom; each distribution plot shows the local

contribution value. The presence of non-measurable tumor lesions resulted in a high SHAP value (0.31), whereas low feature values were related to high SHAP values. Thus, a lower presence of non-measurable tumor lesions would indicate better anti-PD-1 therapy. The NLR showed a high SHAP value (0.32), whereas a lower NLR was related to a better anti-PD-1 response. The PD-L1 percentage was also related to the SHAP value (0.32), and higher PD-L1 percentages indicated better anti-PD-1 responses (**Figure 5A**). The results of hierarchical clustering based on the correlations of feature contribution, PD-L1 percentage, immunotherapy type, smoking dosage, and pathology demonstrated positive correlations with anti-PD-1 therapy, whereas the other factors showed negative correlations with anti-PD-1 therapy (**Figure 5B**). Significant correlations between each selected clinical feature were not observed **(Figure 5C)**.

Figure 5. SHAP values and feature interaction scores in LightGBM-based prediction. (A) SHAP values of individual features were measured. The key color indicates the feature value (high, yellow; low, blue). The SHAP value indicates the response to anti-PD-1 treatment. **(B)** Hierarchical clustering of feature correlation. **(C)** Heatmap of feature correlation (high, yellow; low, blue).

Validation of the performance of LightGBM and XGBoost models in an independent set

The LightGBM and XGBoost models were chosen as the optimal models to predict anti-PD-1 therapy response using the discovery set. We next verified their performance using an independent validation set, which was composed of 26 responders and 24 non-responders and showed a long duration of response in responders (**Figure 6A**). The median PFS and OS in total patients in the validation set

were 5.3 and 13.2 months, respectively. The cut-off values (0.452 in the LightGBM model and 0.476 in the XGBoost model) determined in the discovery set were then applied to evaluate the predictive performance of each model (**Figure 6B, C**). The PPV was 81.4% (22/27) and 81.4% (22/27) and the NPV was 82.6% (19/23) and 82.6% (19/23) in LightGBM and XGBoost models, respectively. The patients with a cut-off value ≥ 0.452 in LightGBM and ≥ 0.476 in XGBoost presented a significantly longer median PFS than the others (10.9 vs. 1.8 months, $P < 0.01$ in LightGBM, 10.9 vs. 1.84 months $P < 0.01$ in XGBoost) (**Figure 6D**), suggesting these models predicted PFS in response to anti-PD-1 therapy in the validation set.

Figure 6. Validation of machine learning-based prediction algorithms for anti-PD-1-treated patients with NSCLC. The validation set of patients with NSCLC was analyzed using the two best machine learning algorithms, LightGBM and XGBoost. **(A)** Swimmer's plot of anti-PD-1-treated patients with advanced NSCLC (stage 4; n=50). The left heatmap shows the results of RECIST (red, partial response; gray, stable disease; blue, progressive disease), and the right heatmap indicates RECIST-based responses (orange, responder; black, non-responder). The colors of the lines and symbols are labeled in the symbol index (lower bottom index). **(B)** Scores of the LightGBM model and statistical analysis. **(C)** Score of the XGBoost model and statistical analysis **(D)** Kaplan-Meier plot of anti-PD-1-treated patients with NSCLC using LightGBM and XGBoost model-based prediction in the validation set. The red line indicates the predicted responder group and the black line indicates the non-responder group. The left plot is for LightGBM and the right plot is for XGBoost. The log-rank methods were used for statistical analyses between responder and non-responder groups (**$p < 0.01$).

Discussion

IHC-based PD-L1 expression is the only approved predictive marker for anti-PD-1 therapy, but its accuracy is not sufficient to discriminate the responders from non-responders. Recently, the tissue and blood tumor mutational burden was also assessed as a potential biomarker independent of PD-L1 expression. However, like PD-L1, the results for tumor mutational burden may vary according to different sequencing platforms, and testing costs are high for routine clinical practice. Therefore, the accuracy of the available biomarkers is still insufficient for deciding whether to treat patients with anti-PD-1 therapy.

The associations of various clinical factors or blood test values in routine practice with treatment outcomes following anti-PD-1 therapy are increasingly being reported. However, the practical clinical application of these findings is limited because these clinical factors and blood test results have been sporadically reported and only simple statistical analyses such as univariate or multivariate have been produced. To establish a predictive model for anti-PD-1 therapy with comprehensive integration of relevant clinical factors, we collected various clinical features from NSCLC patients treated with anti-PD-1 therapy. We selected 19 clinically relevant features from EMRs based on our knowledge, including PD-L1 expression, tumor burden, NLR, smoking years, previous line of therapy, ECOG, and the presence of brain, liver, bone, and adrenal metastases. These factors are evaluated in routine clinical practice and do not require additional cost or efforts. The performance of each clinical feature was tested in the discovery set. The top three features to predict an anti-PD-1 response were PD-L1 expression, tumor burden, and NLR. For comprehensive integration of clinical features and correlative analysis with treatment outcomes, we used ML platforms. Nine ML platforms were tested using these selected 19 clinical features, and the

LightGBM and XGBoost models were chosen to best predict the anti-PD-1 response. Compared to the prediction performance of each single clinical feature, ML based on LightGBM and XGBoost using the combination of clinical features showed more significant performance. Both models showed higher significant prediction accuracy by cut-off value. The patients with values above the cut-off of the models presented a significantly longer median PFS than those with values below the cut-off. Therefore, the cut-off level for these models could discriminate PFS in response to anti-PD-1 therapy.

ML is widely used in many areas within the healthcare industry, from diagnosis and prognosis to drug development, and has significant potential to transform the medical landscape [28]. Indeed, there is a growing realization of the potential of ML as a platform that can integrate information from numerous sources for improvement of decision-making processes for highly skilled workers [28]. Therefore, the potential to utilize ML to aggregate large datasets could significantly accelerate the process of disease identification. Several recent studies have used ML methods to predict immunogenicity and response to certain therapies. For example, Bao et al. [29]. reported the immune landscape and a novel immunotherapy-related gene signature associated with better clinical outcomes in early-stage lung adenocarcinoma using an ML method. Additionally, Duhaze et al. [30]. reported the prediction of immunogenicity for biotherapies using patient- and drug-related factors with massive amounts of data from ML algorithms. To the best of our knowledge, this is the first study to apply ML to establish a clinical decision support algorithm for anti-PD-1 therapy in NSCLC patients.

The LightGBM and XGBoost models selected from the nine ML methods tested here are ensemble-based boosting algorithms. LightGBM is a gradient boosting framework that uses a tree-based learning algorithm. LightGBM can handle large

datasets and takes less memory to run, yielding highly accurate results. This algorithm also supports GPU learning; therefore, data scientists are widely using this model for developing various data science applications. Here, we also applied a random forest model and confirmed the predictive ability with the same discovery set. This was as robust as our LightGBM or XGBoost model; however, the prediction ability was significantly reduced when a new validation set was added (data not shown).

Although they use the same ensemble model, LightGBM and XGBoost are both able to reduce variance and bias. Additionally, this type of ML model trains a new type of model from a preconstructed model. The boosting algorithm employs AdaBoost and gradient boosting methods. Owing to the high analytical efficiency of gradient enhancement technology, its use exceeds that of AdaBoost. However, the limitations of the boosting method are related to speed and overfitting. To overcome this problem, the XGBoost model was developed. XGBoost can compensate for various problems such as overfitting and prediction power, but is still limited in terms of running time. LightGBM can handle large datasets and requires less memory to obtain highly accurate results.

Here, these two models showed comparable predictability. In the clinic, we recommend using LightGBM, which is sufficiently robust to analyze datasets, even with lower computational resources. ML models recognize complex patterns in the data and model the degree to which each feature affects the prediction for each pattern. Therefore, they perform better than typical linear models when features are not independent.

Overall, we obtained more accurate results using LightGBM and XGBoost than using PD-L1 alone by employing data collected noninvasively to stratify patients according to feature patterns and by using nonlinear models to make predictions

based on each pattern. These models, based on clinical features, can serve as useful clinical decision-supporting algorithms that predict anti-PD-1 response in NSCLC patients.

Acknowledgements

Availability of data and material

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no conflicts of interest that are relevant to this study

Funding

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation funded by the Ministry of Science and ICT [grant numbers NRF-2019M3A9B6065231, 2017M3A9E802971, 2017M3A9E9072669 to H.R.K].

Authors' contributions

Kyoung-Ho Pyo: Conceptualization, Methodology, Resources, Investigation, Formal analysis, Roles/Writing - original draft, Writing - review & editing **Beung-Chul Ahn:** Conceptualization**,** Methodology, Resources, Investigation, Formal analysis, Roles/Writing - original draft, Writing - review & editing **Jea-Woo So:** Conceptualization**,** Methodology, Resources, Investigation, Formal analysis, Roles/Writing - original draft, Writing - review & editing **Byoung Chul Cho:**

Conceptualization**,** Methodology, Funding acquisition, Project administration, Resources, Roles/Writing - original draft, Writing - review & editing **Hye Ryun Kim:** Conceptualization**,** Methodology, Funding acquisition, Project administration, Resources, Roles/Writing - original draft, Writing - review & editing **Tae Hyoung Kim:** Project administration, Roles/Writing - original draft, Writing - review & editing **Chang Gon Kim:** Resources, Investigation, Roles/Writing - original draft, Writing - review & editing **Sun Min Lim:** Resources, Investigation, Roles/Writing - original draft, Writing - review & editing **Min Hee Hong:** Resources, Investigation, Roles/Writing - original draft, Writing - review & editing

References

1. Pardoll, D. M. (2012). "The blockade of immune checkpoints in cancer immunotherapy." Nature Reviews Cancer **12**(4): 252-264.
2. Herbst, R. S., et al. (2014). "Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients." Nature **515**(7528): 563-567.
3. Karlsson, A. K. and S. N. Saleh (2017). "Checkpoint inhibitors for malignant melanoma: a systematic review and meta-analysis." Clin Cosmet Investig Dermatol **10**: 325-339.
4. Liu, B., et al. (2017). "Recent development in clinical applications of PD-1 and PD-L1 antibodies for cancer immunotherapy." J Hematol Oncol **10**(1): 174.
5. Garon, E. B., et al. (2015). "Pembrolizumab for the treatment of non-small-cell lung cancer." N Engl J Med **372**(21): 2018-2028.
6. Bagley, S. J., et al. (2017). "Pretreatment neutrophil-to-lymphocyte ratio as a marker of outcomes in nivolumab-treated patients with advanced non-small-cell lung cancer." Lung Cancer **106**: 1-7.

7. Diem, S., et al. (2017). "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 **111**: 176-181.

8. Li, B., et al. (2018). "Impact of smoking on efficacy of PD-1/PD-L1 inhibitors in non-small cell lung cancer patients: a meta-analysis." Onco Targets Ther **11**: 3691-3696.

9. Pantano, F., et al. (2020). "Prognostic clinical factors in patients affected by non-small-cell lung cancer receiving Nivolumab." Expert Opinion on Biological Therapy **20**(3): 319-326.

10. Conforti, F., et al. (2018). "Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis." The Lancet Oncology **19**(6): 737-746.

11. Funazo, T., et al. (2017). "Liver Metastasis Is Associated with Poor Progression-Free Survival in Patients with Non–Small Cell Lung Cancer Treated with Nivolumab." Journal of Thoracic Oncology **12**(9): e140-e141.

12. Shiroyama, T., et al. (2018). "Clinical Characteristics of Liver Metastasis in Nivolumab-treated Patients with Non-small Cell Lung Cancer." Anticancer Res **38**(8): 4723-4729.

13. Garassino, M. C., et al. (2018). "Italian Nivolumab Expanded Access Program in Nonsquamous Non-Small Cell Lung Cancer Patients: Results in Never-Smokers and EGFR-Mutant Patients." J Thorac Oncol **13**(8): 1146-1155.

14. Svaton, M., et al. (2018). "Chronic Inflammation as a Potential Predictive Factor of Nivolumab Therapy in Non-small Cell Lung Cancer." Anticancer Res **38**(12): 6771-6782.

15. Wiesweg, M., et al. (2019). "Machine learning-based predictors for immune checkpoint inhibitor therapy of non-small-cell lung cancer." Annals of Oncology **30**(4): 655-657.

16. Heo, J., et al. (2019). "Machine Learning–Based Model for Prediction of Outcomes in Acute Stroke." Stroke **50**(5): 1263-1265.

17. Cruz, J. A. and D. S. Wishart (2006). "Applications of Machine Learning in Cancer Prediction and Prognosis." Cancer Informatics **2**: 117693510600200030.

18. Kourou, K., et al. (2015). "Machine learning applications in cancer prognosis and prediction." Computational and Structural Biotechnology Journal **13**: 8-17.

19. Rech, A. J., et al. (2018). "Radiotherapy and CD40 Activation Separately Augment Immunity to Checkpoint Blockade in Cancer." Cancer Res **78**(15): 4282-4291.

20. Chen, T. and C. Guestrin (2016) XGBoost: A Scalable Tree Boosting System. arXiv e-prints arXiv:1603.02754

21. Ke, G., et al. (2017). "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." 3146--3154.

22. Krogh, A. (2008). "What are artificial neural networks?" Nat Biotechnol **26**(2): 195-197.

23. Lee, Y. (2010). "Support vector machines for classification: a statistical portrait." Methods Mol Biol **620**: 347-368.

24. Abello, J. and G. Cormode (2007). Discrete Methods in Epidemiology, American Mathematical Society.

25. Ryback, R. S., et al. (1982). "Quadratic discriminant analysis as an aid to interpretive reporting of clinical laboratory tests." Jama **248**(18): 2342-2345.

26. Gogtay, N. J., et al. (2017). "Principles of Regression Analysis." J Assoc Physicians India **65**(4): 48-52.

27. Lundberg, S. and S.-I. Lee (2017) A Unified Approach to Interpreting Model Predictions. arXiv e-prints arXiv:1705.07874

28. (2019). "Ascent of machine learning in medicine." Nature Materials **18**(5): 407-407.

29. Bao, X., et al. (2020). "Immune landscape and a novel immunotherapy-related gene signature associated with clinical outcome in early-stage lung adenocarcinoma." J Mol Med (Berl) **98**(6): 805-818.

30. Duhazé, J., et al. (2020). "A Machine Learning Approach for High-Dimensional Time-to-Event Prediction With Application to Immunogenicity of Biotherapies in the ABIRISK Cohort." Front Immunol **11**: 608.