ARTIFICIAL INTELLIGENCE

original reports abstract

Dynamic Risk Prediction of 30-Day Mortality in Patients With Advanced Lung Cancer: Comparing Five Machine Learning Approaches

Charles Vesteghem, PhD[1,2,3]; Weronika M. Szejniuk, MD, PhD[1,3,4]; Rasmus F. Brøndum, PhD[1,2,3]; Ursula G. Falkmer, MD, PhD[1,3,4]; Chloé-Agathe Azencott, PhD[5,6,7]; and Martin Bøgsted, PhD[1,2,3]

PURPOSE Administering systemic anticancer treatment (SACT) to patients near death can negatively affect their health-related quality of life. Late SACT administrations should be avoided in these cases. Machine learning techniques could be used to build decision support tools leveraging registry data for clinicians to limit late SACT administration.

MATERIALS AND METHODS Patients with advanced lung cancer who were treated at the Department of Oncology, Aalborg University Hospital and died between 2010 and 2019 were included (N = 2,368). Diagnoses, treatments, biochemical data, and histopathologic results were used to train predictive models of 30-day mortality using logistic regression with elastic net penalty, random forest, gradient tree boosting, multilayer perceptron, and long short-term memory network. The importance of the variables and the clinical utility of the models were evaluated.

RESULTS The random forest and gradient tree boosting models outperformed other models, whereas the artificial neural network–based models underperformed. Adding summary variables had a modest effect on performance with an increase in average precision from 0.500 to 0.505 and from 0.498 to 0.509 for the gradient tree boosting and random forest models, respectively. Biochemical results alone contained most of the information with a limited degradation of the performances when fitting models with only these variables. The utility analysis showed that by applying a simple threshold to the predicted risk of 30-day mortality, 40% of late SACT administrations could have been prevented at the cost of 2% of patients stopping their treatment 90 days before death.

CONCLUSION This study demonstrates the potential of a decision support tool to limit late SACT administration in patients with cancer. Further work is warranted to refine the model, build an easy-to-use prototype, and conduct a prospective validation study.

JCO Clin Cancer Inform 6:e2200054. © 2022 by American Society of Clinical Oncology

PURPOSE

Systemic anticancer treatments (SACTs) includes chemotherapy, targeted therapy, immunotherapy, and hormonal therapy treatments. A SACT should only be considered in patients with an adequate benefit from the treatment since SACTs often have a short-term negative impact on health-related quality of life.[1-7] An accepted threshold for late SACT administration is 30 days before death.[8] However, clinicians' experience in predicting the remaining lifetime of patients may be inadequate,[9] leading to prescription of SACT too late to achieve a clinical benefit.[8] Furthermore, death from advanced cancer often has a multifactorial background where acute complications could lead to patient death.

Lung cancer is a frequently occurring cancer type with poor prognosis and high mortality, particularly in advanced stages. Thus, patients with lung cancer are at higher risk of receiving SACT close to death than other cancer types with a better prognosis.

There is a need for decision support tools to assist the work of clinicians to minimize the risk of decreasing health-related quality of life because of SACT of patients with lung cancer receiving palliative treatment in advanced stages. Patient health might promptly deteriorate during treatment, requiring frequent use of dynamic predictive tools to assess their situation. To the best of our knowledge, existing studies addressing this issue (1) are based on a limited number of clinical variables; (2) do not consider artificial neural network–based models; (3) are based on different end points, for example, 6-month mortality; or (4) are not suitable for dynamic risk prediction.[10-17]

The aim of this study was to investigate the potential use of machine learning approaches on electronic

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on September 29, 2022 and published at ascopubs.org/journal/cci on November 15, 2022: DOI https://doi.org/10.1200/CCI.22.00054

CONTEXT

Key Objective

How can machine learning help clinicians limit the administration of systemic anticancer treatment (SACT) to patients near death?

Knowledge Generated

Across the five machine learning models tested, the gradient boosting and random forest models performed the best and demonstrated their utility by identifying 40% of late SACT, defined as more than 50% probability of 30-day mortality at the time of administration, whereas neural network–based models underperformed. Biochemical results contained most of the information in the prediction.

Relevance

This work paves the way for implementation of decision support tools to help clinicians assess the relevance of palliative SACT administration in lung cancer.

health registers and administrative data to limit late SACT by building dynamic predictive models for the 30-day mortality of patients with advanced lung cancer. It is based on the hypothesis that extensive medical data can improve the performances of the predictive models and on the hypothesis that artificial neural network–based machine learning techniques can outperform other methods.

MATERIALS AND METHODS

Study Population

Patients in contact with the Department of Oncology between January 1, 2008, and December 31, 2019, and who died between January 1, 2010, and December 31, 2019, were identified from the patient administrative system (n = 14,902). Among these patients, 3,856 were diagnosed with lung cancer. Only those who received SACT for advanced or metastatic lung cancer or who were diagnosed with metastatic lung cancer were included in the final data set (N = 2,368, Data Supplement). The patients were split into three cohorts: the training cohort with patients who died between 2010 and 2017, the validation cohort with patients who died in 2018, and the test cohort for patients who died in 2019. Since only patients in contact with the Department of Oncology between 2008 and 2019 were accessible, including patients who died in the same period would exclude patients who died in this period but were in contact with the Department of Oncology only before 2008. To avoid missing this type of patient, only patients who died after 2010 were included. A 2-year margin was considered sufficient in this context. Using the date of death instead of the date of diagnosis allows us to limit the class imbalance in the recent cohorts since in the latter case, the records close to death would not be available for a large proportion of patients.

Data Sources and Variables

This study was based on five data sources from the North Denmark Region and included demographic data, cancer diagnoses, comorbidities, symptoms and side effects, surgeries, radiotherapies, drug prescriptions and administrations, BMI, biochemical data, morphological data, and biomarker data (see the Data Supplement, Data sources and variables). The data were merged at the patient level with the Danish civil registration number and were subsequently pseudonymized.[18] The data set contained three different types of variables: baseline, cumulative, and status. Baseline variables represented information up to 30 days before diagnosis and were constant for each patient across the sequence of records (Fig 1). Status variables represented a potentially variable state, such as BMI, biochemical data, diagnoses, or biomarkers. Cumulative variables included those that could be counted or summed and were treatment-related, eg, the number of a certain type of surgery or cumulative dose of a certain drug. For status variables, the last value was carried forward in the case of missing values, whereas for the cumulative variables, empty values were designated with zeros. If no value was still available, the mean value was used.

In addition to these variables, referred to as base variables, summary variables were generated for all cumulative and status variables. For cumulative variables, the sums of values from diagnosis to the current record and in the past 30 days[17] were computed. For status variables, the difference with the mean of all previous values was generated.

Data management was performed using SAS Enterprise Guide 8.3 (SAS Institute Inc, Cary, NC) and Python 3.8 with Jupyter[18] notebooks.

Data Set Generation

The overall data set included records only after the start of the first palliative treatment or, if no palliative treatment was initiated, the first diagnosis of metastatic disease. Only records or sequences of records associated with contact with the Department of Oncology were retained.

Two versions of the data set were created, one with only the base variables, referred to as the base data set, and the other

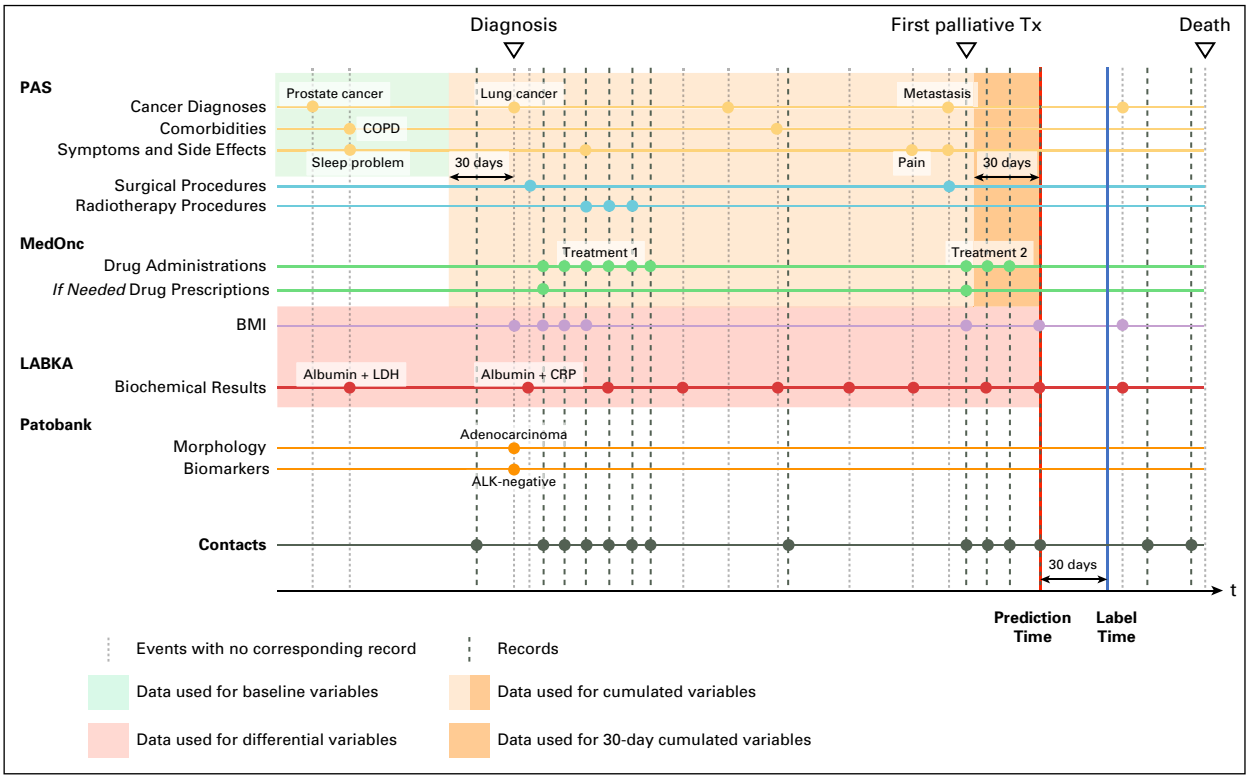

FIG 1. Example of records for a patient. The prediction time corresponds to the time point where the prediction is made, with the prediction being whether the patient is dead or alive at the label time. Each dot on the figure represents a data point. ALK, anaplastic lymphoma kinase gene; BMI, body mass index; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; LDH, lactate dehydrogenase; Tx, treatment.

with both the base and summary variables (see the Data Supplement, Data set generation and feature selection).

In a second phase, an additional data set, referred to as the biochemical results data set, was generated keeping only the biochemical variables to compare performance with the base data set. Two versions of this data set were also created and filtered as mentioned above.

Models

Five popular machine learning models were considered in this study: a logistic regression with elastic net regularization (LRENR),[19] a random forest classifier (RF),[20] a gradient tree boosting classifier (GB),[21] a multilayer perceptron (MLP), and a long short-term memory model (LSTM)[22] with a basic architecture, as proposed in the literature[13,14,23] (see the Data Supplement, Models). A simple logistic regression model was used as baseline. All models were trained using the training set with various values for the hyperparameters, and the performance was evaluated on the validation set to select the best set of hyperparameters. To maintain the temporal structure of the data, no cross-validation was performed. The final performance was evaluated on the records from the held-out test cohort.[24]

To assess the variability of the performance for all models, nine additional training sets of the same length were generated by bootstrapping the combined training and validation records. The significance of the difference in performance was evaluated by the Wilcoxon signed-rank test. The corresponding *P* values were calculated using the Python SciPy library.[25] A threshold of 0.05 for *P* values was used for significance.

LRENR, RF, and GB were fitted using the Python scikit-learn library,[26] and hyperparameters were optimized using a grid search. The two artificial neural network–based models (MLP and LSTM) were trained using the Python Keras library[27] and optimized by Bayesian optimization using the Python Keras-Tuner library.[28]

The importance of each variable was estimated from their Shapley Additive explanation (SHAP) values[29] on a random sample of 1,000 records from the combined training and validation records.

Evaluation of Utility

To assess the usefulness of a predictive model, the potential effect of limiting late SACT administrations on the basis of a simple rule was investigated. Given a threshold on the 30-day mortality risk score, SACT should be administered if the predicted risk is below that threshold. Conversely, if the predicted risk is above the threshold, SACT should not be given. In cases where the risk is above the threshold at a given time point but becomes below the threshold at a later stage, SACT is considered administrable at that later time

point and is therefore only considered delayed. An administration was considered preventable if the risk prediction at the time of administration and for all subsequent contacts were above the threshold (Data Supplement).

To produce interpretable risk predictions, the considered models were calibrated by isotonic regression, using the Python scikit-learn library.[30] Risk was predicted at each drug administration event in the test cohort records. A 50% threshold-based rule was applied to these predictions to identify preventable SACT administrations within 30 days and 90 days and more than 90 days from death and the corresponding number of patients.

Ethical Approval and Registration

The study was registered at the North Denmark Region's research project inventory (reg. No. 2019-41) and approved by the Danish Patient Safety Authority (case No. 31-1521-334).

RESULTS

Study Population

The population characteristics are described in Table 1. The overall cohort was well balanced regarding sex (52% male and 48% female). A majority of patients (56%) died between age 60 and 74 years, and most patients (87%) received palliative SACT, among whom 65% died $<$ 12 months from the initiation of palliative treatment. The most prevalent histopathologic subtype was adenocarcinoma (43%), followed by small cell carcinoma (25%). An increasing proportion of adenocarcinoma and decreasing incidence of small-cell lung cancer were observed between the training, validation, and test cohorts. The number of patients surviving more than 12 months was increased in the test (50.5%) cohort compared with the training (32.6%) and validation cohorts (35.4%).

TABLE 1. Study Population Characteristics

Category	Training Set, No. (%)	Validation Set, No. (%)	Test Set, No. (%)	Overall, No. (%)
Patients	1,819 (100.0)	309 (100.0)	240 (100.0)	2,368 (100.0)
Sex				
Male	937 (51.5)	168 (54.4)	117 (48.8)	1,222 (51.6)
Female	882 (48.5)	141 (45.6)	123 (51.3)	1,146 (48.4)
Histopathology				
Adenocarcinoma	753 (41.4)	143 (46.3)	122 (50.8)	1,018 (43.0)
Small-cell carcinoma	493 (27.1)	64 (20.7)	43 (17.9)	600 (25.3)
Large-cell carcinoma	264 (14.5)	30 (9.7)	27 (11.3)	321 (13.6)
Squamous cell carcinoma	223 (12.3)	57 (18.4)	40 (16.7)	320 (13.5)
Others	86 (4.7)	15 (4.9)	8 (3.3)	109 (4.6)
Age at death, years				
18-44	35 (1.9)	3 (1.0)	5 (2.1)	43 (1.8)
45-59	416 (22.9)	54 (17.5)	47 (19.6)	517 (21.8)
60-74	1,071 (58.9)	165 (53.4)	132 (55.0)	1,368 (57.8)
$\geq$ 75	297 (16.3)	87 (28.2)	56 (23.3)	440 (18.6)
Palliative treatment				
Yes	1,657 (91.1)	226 (73.1)	194 (80.8)	2,077 (87.7)
No	162 (8.9)	83 (26.9)	46 (19.2)	291 (12.3)
Survival from the start of palliative treatment, months				
0-1	105 (6.3)	9 (4.0)	9 (4.6)	123 (5.9)
1-6	526 (31.7)	70 (31.0)	46 (23.7)	642 (30.9)
6-12	486 (29.3)	67 (29.6)	41 (21.1)	594 (28.6)
$\geq$ 12	540 (32.6)	80 (35.4)	98 (50.5)	718 (34.6)
Contacts				
All contacts	68,876 (100.0)	12,205 (100.0)	12,209 (100.0)	93,290 (100.0)
Within 30 days of death	10,783 (15.7)	2,190 (17.9)	1,526 (12.5)	14,499 (15.5)

NOTE. The percentages in parentheses present the proportion of corresponding patients in the cohort, except for the survival data, where the value is the proportion of corresponding patients among patients who received palliative treatment, and the contact data, which represent the proportion of contacts.

Performances

Comparing models. The average precision (AP) and receiver operating characteristic area under the curve (ROC AUC) of the five models were computed for all data sets (Fig 2). First, all values from the base data set, with or without summary variables, were considered. There were limited differences in the validation set, with mean values for the AP varying between 0.486 and 0.544. The inclusion of summary variables had a beneficial effect on the performance. The differences were larger on the test set where values between 0.342 and 0.509 were observed, with the MLP and LSTM underperforming compared with other models ($P < .002$). The effect of the summary variables was beneficial for the performances of all applicable models except the MLP model but was not significant for the GB and RF models, changing from 0.500 to 0.505 ($P = .232$) and from 0.498 to 0.509 ($P = .097$), respectively.

The patterns were similar using the ROC AUC as a performance metric, with values between 0.794 and 0.846 in the validation set and between 0.766 and 0.868 in the test set, with the MLP and LSTM models performing poorly on the test set ($P < .002$).

Performances using only biochemical results. Most of the important variables were biochemical results, notably albumin, leukocytes, carbamide, and lactate dehydrogenase (Data Supplement). The performance of models trained exclusively on these variables was therefore investigated (Fig 2). This had a limited but significant negative impact on the performance of RF and GB. The mean AP values

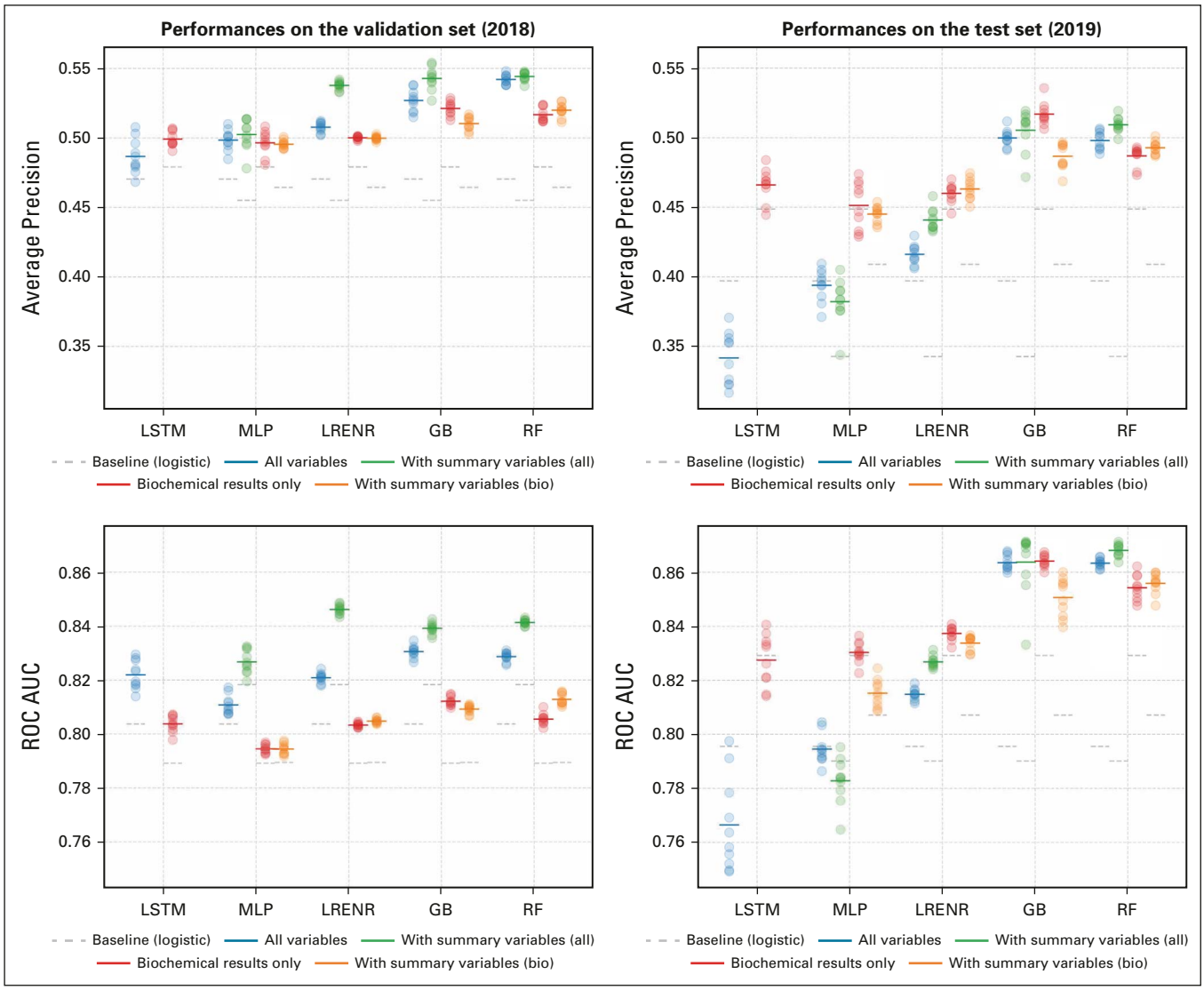

FIG 2. Average precision and ROC AUC per model and data set on the validation and test sets. The baseline values show the performance of a logistic regression without regularization. The horizontal lines represent the mean value of the corresponding metric. Each circle represents the performance on a bootstrapped data set. GB, gradient tree boosting classifier; LRENR, logistic regression with elastic net regularization; LSTM, long short-term memory model; MLP, multilayer perceptron; RF, random forest classifier; ROC AUC, receiver operating characteristic area under the curve.

between the base data set and the biochemical results data set, both with the summary variables, went from 0.509 to 0.493 ($P < .004$) and from 0.505 to 0.487 ($P < .002$) in the RF and GB models, respectively. Notably, the optimum performance for the GB model was obtained for the biochemical results data set without the summary variables with a mean AP of 0.517. This was highly significantly better than for any other combinations of models and data sets ($P < .002$), except when compared with the RF and GB models on the base data set with summary variables ($P = .048$ and $P = .064$, respectively).

Utility

The GB model was chosen as the overall best model, in terms of AP using the biochemical results data set without summary variables, to assess the utility of a decision support tool in preventing late SACT administrations.

In the test cohort of the 195 patients who received palliative SACT, 16% (n = 32) received, on average, 3.2 (103/32) administrations within 30 days of death (Table 2).

Using the 50% threshold, 31% (10 of 32) of patients from the test cohort could have had preventable SACT administrations within 30 days of death, corresponding to 40% (41 of 103) of the late SACT administrations (Fig 3). However, this threshold led to SACT administrations that would have been considered as preventable before the 90-day landmark for 2% (4 of 195) of patients. The 41 preventable late SACT administrations were primarily for osimertinib, etoposide, alectinib, and carboplatin.

The F_1 scores for the GB and RF models and all data sets were calculated at both the patient and administration levels in Table 3 with heterogeneous results.

DISCUSSION

Five different machine learning models were compared for dynamic prediction of 30-day mortality in patients with advanced and metastatic lung cancer, including two artificial neural network–based models and two tree-based ensemble models. The two tree-based ensemble models performed best and exhibited similar performances in terms of average precision. When used on the final test set, for the two models RF and GB, the ROC AUC increased, whereas the negative impact on AP was limited.

The performances of the models were in most cases only marginally affected, even increasing in some cases, when

TABLE 2. Utility Evaluation for the Prevention of SACT Administration Within 30 days of Death in the Gradient Tree Boosting Classifier Model on the Biochemical Results Data Set Without Summary Variables

	Administered		Preventable						
Category	Total	Within 30 Days	Total	Within 30 Days	Within 90 Days	Before 90 Days	Precision	Recall	F1 Score
Treated patients	195	32 (16.4%)	20 (10.3%)	10 (5.1%)	6 (3.1%)	4 (2.1%)	0.7	0.3	0.4
Administrations	4,154	103 (2.5%)	77 (1.9%)	41 (1.0%)	28 (0.7%)	8 (0.2%)	0.8	0.4	0.5
Etoposide	500-1,000	4%	3%	1%	2%	0%	1.0	0.3	0.5
Vinorelbine	500-1,000	1%	1%	0%	1%	0%	0.4	0.3	0.3
Carboplatin	500-1,000	3%	2%	1%	1%	0%	0.8	0.3	0.5
Nivolumab	200-500	1%	0%	0%	0%	0%	NA	0.0	NA
Pemetrexed	200-500	2%	1%	0%	0%	0%	0.0	0.0	NA
Osimertinib	200-500	8%	5%	5%	0%	0%	1.0	0.6	0.8
Gemcitabine	50-200	0%	1%	0%	1%	0%	NA	NA	NA
Docetaxel	50-200	2%	2%	1%	1%	0%	1.0	0.7	0.8
Pembrolizumab	50-200	1%	2%	1%	0%	1%	0.3	0.5	0.4
Topotecan	50-200	8%	0%	0%	0%	0%	NA	0.0	NA
Alectinib	50-200	14%	14%	14%	0%	0%	1.0	1.0	1.0

NOTE. The Administered columns inform on the number of SACT administrations given. The corresponding Within 30 Days column informs on the number of SACT administrations given within 30 days of death. The Preventable column is for the numbers of SACT administrations that would have been prevented using the threshold as mentioned in Section 2.4. The corresponding Within 30 Days column informs on the number of SACT drug administrations that would have been preventable within 30 days of death. The Within 90 Days column provides the number of SACT drug administrations that would have been preventable between 90 days and 30 days to death. The Before 90 Days column informs on the number of SACT drug administrations that would have been preventable more than 90 days from death. The Treated patients row provides information on the number of corresponding patients for each column. For individual drugs, only the range of total administrations is shown, and the percentages shown are in relation to the total number of administrations or patients and are rounded to the closest percent for anonymization.

Abbreviation: SACT, systemic anticancer treatment.

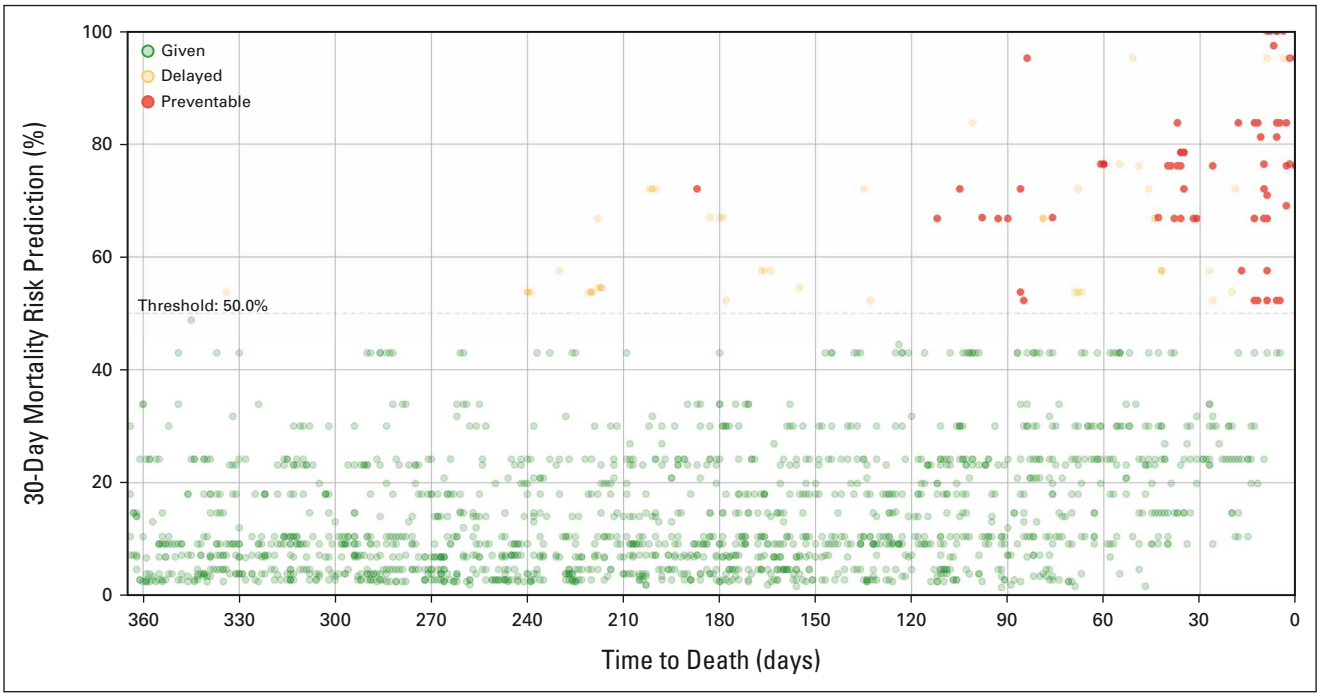

FIG 3. Thirty-day mortality risk predictions in the past 365 days before death for patients included in the test data set using the RF model with summary variables. Each circle represents a SACT drug administration. The trajectories are not represented, and the days to death values have been randomly shifted within 2 days of the actual values for anonymization. RF, random forest classifier; SACT, systemic anticancer treatment.

using only the biochemical results. Inclusion of summary variables had no clear benefit.

A utility analysis was performed on the tree-based ensemble models, indicating that some late SACT administrations could be prevented by implementing the predictive model with a simple threshold-based rule while maintaining almost all treatments before the 90-day landmark.

One of the main strengths of this study is the long inclusion timeframe of patients, allowing us to thoroughly investigate temporal performance of the models.

In addition, the single center study guaranties a high level of consistency and allows us to access detailed clinical information. On the one hand, this leads to questions regarding the generalizability of these results outside this center since treatment implementations might vary among hospitals. On the other hand, there is high homogeneity in the Danish health care system, which should allow an extension of these results to the entire country.

Biochemical results are potentially highly variable and capture information that can change significantly between the periods before and after the 30-day threshold. They seem to contain most of the information needed for accurate prediction, notably that drug administrations do not seem to have a real impact. A simpler model using only these data should therefore be encouraged. It also indicates that these results can be broadly extended to other institutions, because they are less dependent on local practices, being less subject to interpretation, or to other diagnoses sharing the same type of progression, such as pancreatic cancer. This is consistent with many studies investigating the prediction of mortality in patients with cancer, which primarily use biochemical results in their models.[11,14,15,31]

Another key aspect is the validity of the model over time. Clinical practice changes over the years, notably with the introduction of new treatment options and better management of side effects so that a model fitted on old data might not be applicable to current data of patients with lung cancer. This was the reasoning for using patients who died in 2018 and 2019 as validation and test sets, respectively.

Inclusion of summary variables to inform patient trajectories yielded mixed results. Some models benefitted from it, notably the RF model, whereas it had varying impacts on other models, such as the GB model. Not including these variables could prevent overfitting in some cases. The same mechanism was potentially observed for nonbiochemical variables that could cause overfitting. An important aspect is how well these summary variables inform the trajectory of the patients. Indeed, each variable probably exhibits different dynamics near the end of life, which would require the design of specific summary variables.

Studies in similar contexts have also reported good performance of gradient boosting and the often-poor performances of artificial neural network–based models trained on tabular data.[14,17,23] Indeed, artificial neural network–based models are difficult to tune because of the number of

TABLE 3. Comparison of the RF and GB Models on All Four Data Sets in Terms of F1 Score for Utility Evaluation

Model and Data Set	Patients				Administered			
	Within 30 Days, No. (%)	Within 90 Days, No. (%)	Before 90 Days, No. (%)	F1 Score	Within 30 Days, No. (%)	Within 90 Days, No. (%)	Before 90 Days, No. (%)	F1 Score
RF								
Base data set	14 (7.2)	17 (8.7)	15 (7.7)	0.46	67 (1.6)	76 (1.8)	65 (1.6)	0.57
With summary variables	15 (7.7)	17 (8.7)	11 (5.6)	0.52	59 (1.4)	53 (1.3)	29 (0.7)	0.62
Biochemical results data set	11 (5.6)	5 (2.6)	5 (2.6)	0.46	43 (1.0)	22 (0.5)	10 (0.2)	0.55
With summary variables (bio)	16 (8.2)	12 (6.2)	9 (4.6)	0.56	54 (1.3)	43 (1.0)	24 (0.6)	0.60
GB								
Base data set	11 (5.6)	7 (3.6)	6 (3.1)	0.45	40 (1.0)	34 (0.8)	13 (0.3)	0.51
With summary variables	11 (5.6)	3 (1.5)	4 (2.1)	0.47	27 (0.6)	7 (0.2)	4 (0.1)	0.40
Biochemical results data set	10 (5.1)	6 (3.1)	4 (2.1)	0.43	41 (1.0)	28 (0.7)	8 (0.2)	0.54
With summary variables (bio)	9 (4.6)	2 (1.0)	4 (2.1)	0.40	33 (0.8)	10 (0.2)	8 (0.2)	0.46

NOTE. In the Administered section referring to administrations, the Within 30 Days column informs on the number of preventable SACT administrations given within 30 days of death. The Within 90 Days column provides the number of SACT drug administrations that would have been preventable between 90 days and 30 days to death. The Before 90 Days column informs on the number of SACT drug administrations that would have been preventable more than 90 days from death. The value in parentheses is the percent of corresponding administrations among the total number of administrations (Table 2). In the Patients column, the presented values inform on the number of corresponding patients in relation to the drug administrations and values in parentheses are the percentage of corresponding patients among treated patients.

Abbreviations: GB, gradient tree boosting classifier; RF, random forest classifier; SACT, systemic anticancer treatment.

hyperparameters and the instability of the optimization procedure and rarely outperform other approaches on structured data.[32] It could, however, be speculated that this is due to simple choices of architecture, notably for the LSTM. Although better performances could have likely been achieved by fine tuning the architecture of the model, our intent was to evaluate the performances of some comparable, popular baseline models from the literature. Extra tuning work could have also been used to perform more extensive feature engineering for the other models, for example, by including interactions, designing specific summary variables that would better represent the trajectory of the patients, or optimizing other hyperparameters such as the learning rate for the GB model. Accordingly, we decided to maintain relatively simple architectures for the LSTM and MLP to allow for a fair comparison to identify in which direction more effort should be placed.

Another reason for the poor performances of the artificial neural network–based models could be their inability to properly learn from heterogeneous data. Indeed, these models are known to perform well in image recognition and language processing applications that use very homogeneous and correlated data, which is rarely the case with tabular data.[33,34] From this study and the ones mentioned above, we do not believe that neural network–based models should be a recommended approach on tabular data, even with more data points than those included in this study (see Contacts in Table 1), unless a clear strategy in terms of architecture is identified.

SHAP values were used to assess the importance of each variable in the predictions from each model. However, the interpretability of these values is limited compared with that of the coefficient in a linear model such as LRENR. The explainability of nonlinear models such as tree-based models or artificial neural networks is the topic of ongoing research.[35] If explainability is of critical importance, LRENR should be prioritized.

The primary aim of this study was to introduce a predictive model capable of supporting clinical decision making to ordinate SACT and thus potentially limit the risk of unnecessary harm. The crucial aspect of the models is their utility in practice. As shown in Table 3, even after calibration, the performances are quite heterogeneous across models and data sets and thus subject to caution. Nevertheless, this study demonstrates that using a simple rule alongside the predictive model could limit the amount of SACT given too close to death. The rule cannot be implemented as is in a clinical context but could guide oncologists in their decision making. Indeed, many parameters should be considered, notably the type of SACT. For example, protein kinase inhibitors such as osimertinib and alectinib typically have milder side effects while potentially avoiding flaring of the tumor, limiting the interest of stopping the treatment, even close to death. Conversely, etoposide and carboplatin often have much more severe side effects; therefore, a predictive model could help limit their use too close to death.

The goal of this study was to investigate the possibility of constructing a decision support tool to avoid administering unnecessary SACTs in patients with lung cancer. Additional work is needed to develop a prototype applicable in a clinical context and to conduct a prospective validation study, notably in terms of adoption by clinicians. In practice, we envision a web server with a live connection to the EHRs and administrative data with a user-friendly web interface where clinicians can acquire an assessment of the risk of individual patient 30-day mortality by explicitly providing an identifier.

Once the solution is validated, long-term support and maintenance will be required to retrain the model to maintain an acceptable level of performance.

In conclusion, prediction of 30-day mortality in patients with advanced lung cancer was most accurate using tree-based machine learning models on EHRs and administrative data. Most of the information was contained in biochemical parameters, limiting the interest of using other data sets for the prediction, such as comorbidities, disease trajectories, or histopathology. Using predictive modeling may potentially help to limit late SACT, reducing the risk of causing unnecessary harm to patients in the late stage of life.

AFFILIATIONS

[1]Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
[2]Department of Haematology, Aalborg University Hospital, Aalborg, Denmark
[3]Clinical Cancer Research Centre, Aalborg University Hospital, Aalborg, Denmark
[4]Department of Oncology, Aalborg University Hospital, Aalborg, Denmark
[5]CBIO Mines ParisTech, PSL Research University, Paris, France
[6]Institut Curie, Paris, France
[7]INSERM U900, Paris, France

CORRESPONDING AUTHOR

Charles Vesteghem, PhD, Department of Clinical Medicine, Aalborg University, Sdr. Skovvej 15, 9000 Aalborg, Denmark; Twitter: @vesteghem; e-mail: charles.vesteghem@rn.dk.

DISCLAIMER

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPORT

Supported by the Department of Oncology, Aalborg University Hospital, North Denmark Region's Health Innovation Foundation No. 2019-046332, and the Danish Cancer Research Fund.

AUTHOR CONTRIBUTIONS

Conception and design: Charles Vesteghem, Rasmus F. Brøndum, Ursula G. Falkmer, Chloé-Agathe Azencott, Martin Bøgsted
Financial support: Ursula G. Falkmer, Martin Bøgsted
Administrative support: Ursula G. Falkmer, Martin Bøgsted
Provision of study materials or patients: Ursula G. Falkmer
Collection and assembly of data: Charles Vesteghem, Ursula G. Falkmer
Data analysis and interpretation: All authors
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Weronika M. Szejniuk
Travel, Accommodations, Expenses: MSD Oncology

Chloé-Agathe Azencott
Research Funding: Janssen Research & Development (Inst), Sancare (Inst)
Patents, Royalties, Other Intellectual Property: Coauthor on a declaration of invention on sensitive and specific determination of DNA methylation profiles (Institut Curie) (Inst)
Uncompensated Relationships: Sancare, Epigene Labs

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

We would like to thank the System Administrator of MedOnc, Annette Juul Madsen, Department of Oncology, and Special Consultant Thomas Mulvad Larsen, Business Intelligence Unit, North Denmark Region, for their help in obtaining and understanding the data needed for this study.

REFERENCES

1. Hanahan D, Weinberg RA: Hallmarks of cancer: The next generation. Cell 144:646-674, 2011
2. Taylor CW, Kirby AM: Cardiac side-effects from breast cancer radiotherapy. Clin Oncol (R Coll Radial) 27:621-629, 2015
3. Kappers MH, van Esch JHM, Sleijfer S, et al: Cardiovascular and renal toxicity during angiogenesis inhibition: Clinical and mechanistic aspects. J Hypertens 27:2297-2309, 2009
4. Pabla N, Dong Z: Curtailing side effects in chemotherapy: A tale of PKCδ in cisplatin treatment. Oncotarget 3:107-111, 2012
5. Davis C: Drugs, cancer and end-of-life care: A case study of pharmaceuticalization? Soc Sci Med 131:207-214, 2015
6. Li T, Mizrahi D, Goldstein D, et al: Chemotherapy and peripheral neuropathy. Neurol Sci 42:4109-4121, 2021
7. Earle CC, Neville BA, Landrum MB, et al: Evaluating claims-based indicators of the intensity of end-of-life cancer care. Int J Qual Health Care 17:505-509, 2005
8. Wallington M, Saxon EB, Bomb M, et al: 30-day mortality after systemic anticancer treatment for breast and lung cancer in England: A population-based, observational study. Lancet Oncol 17:1203-1216, 2016
9. Christakis NA, Lamont EB: Extent and determinants of error in doctors' prognoses in terminally ill patients: Prospective cohort study. Br Med J 320:469-473, 2000
10. Simmons CPL, McMillan DC, McWilliams K, et al: Prognostic tools in patients with advanced cancer: A systematic review. J Pain Symptom Manage 53:962-970.e10, 2017
11. Hamano J, Takeuchi A, Yamaguchi T, et al: A combination of routine laboratory findings and vital signs can predict survival of advanced cancer patients without physician evaluation: A fractional polynomial model. Eur J Cancer 105:50-60, 2018
12. Adelson K, Lee DK, Velji S, et al: Development of imminent mortality predictor for advanced cancer (IMPAC), a tool to predict short-term mortality in hospitalized patients with advanced cancer. JCO Oncol Pract 14:e168-e175, 2018
13. Renfro LA, Goldberg RM, Grothey A, et al: Clinical calculator for early mortality in metastatic colorectal cancer: An analysis of patients from 28 clinical trials in the Aide et Recherche en Cancérologie Digestive Database. J Clin Oncol 35:1929-1937, 2017
14. Parikh RB, Manz C, Chivers C, et al: Machine learning approaches to predict 6-month mortality among patients with cancer. JAMA Netw Open 2:e1915997, 2019
15. Uneno Y, Taneishi K, Kanai M, et al: Development and validation of a set of six adaptable prognosis prediction (SAP) models based on time-series real-world big data analysis for patients with cancer receiving chemotherapy: A multicenter case crossover study. PLoS One 12:e0183291, 2017
16. Sjoquist KM, Renfro LA, Simes RJ, et al: Personalizing survival predictions in advanced colorectal cancer: The ARCAD Nomogram Project. J Natl Cancer Inst 110:638-648, 2017
17. Elfiky AA, Pany MJ, Parikh RB, Obermeyer Z: Development and application of a machine learning approach to assess short-term mortality risk among patients with cancer starting chemotherapy. JAMA Netw Open 1:e180926, 2018
18. Kluyver T, Ragan-Kelley B, Pérez F, et al: Jupyter Notebooks—A publishing format for reproducible computational workflows, in Positioning and Power in Academic Publishing: Players, Agents and Agendas. Proceedings of the 20th International Conference on Electronic Publishing, ELPUB, Göttingen, Germany, 2016, pp 87-90
19. Zou H, Hastie T: Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol 67:301-320, 2005
20. Ho TK Random decision forests. Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, Montreal, Canada, August 14-16, 1995, pp 278–282
21. Friedman JH: Greedy function approximation: A gradient boosting machine. Ann Stat 29:1189-1232, 2001
22. Hochreiter S, Schmidhuber J: Long short-term memory. Neural Comput 9:1735-1780, 1997
23. Lauritsen SM, Kalør ME, Kongsgaard EL, et al: Early detection of sepsis utilizing deep learning on electronic health record event sequences. Artif Intell Med 104:101820, 2020
24. Hastie T, Tibshirani R, Friedman J: The Elements of Statistical Learning. Springer Series in Statistics, Volume 26. New York, NY, Springer, 2009
25. Virtanen P, Gommers R, Oliphant TE, et al: SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nat Methods 17:261-272, 2020

26. Pedregosa F, Varoquaux G, Gramfort A, et al: Scikit-learn: Machine learning in Python. J Mach Learn Res 12:2825-2830, 2011
27. Chollet F: Keras, 2015. https://keras.io/
28. O'Malley T, Bursztein E, Long J, et al: Keras-tuner, 2019. https://github.com/keras-team/keras-tuner
29. Lundberg S, Lee S-I: A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, 2017
30. Niculescu-Mizil A, Caruana R: Predicting good probabilities with supervised learning. Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, August 2005, pp 625-632
31. Yuan Y, Zhu L, Li X, et al: A new prognostic score based on the systemic inflammatory response in patients with inoperable non-small-cell lung cancer. Onco Targets Ther 9:4879-4886, 2016
32. Shwartz-Ziv R, Armon A: Tabular data: Deep learning is not all you need. Inf Fusion 81:84-90, 2022
33. Devlin J, Chang M-W, Lee K, Toutanova K: BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding, arXiv:1810.04805, 2018
34. Shen L, Margolies LR, Rothstein JH, et al: Deep learning to improve breast cancer detection on screening mammography. Sci Rep 9:12495, 2019
35. Belle V, Papantonis I: Principles and practice of explainable machine learning. Front Big Data 4:688969, 2021