Predicting objective response rate (ORR) in immune checkpoint inhibitor (ICI) therapies with machine learning (ML) by combining clinical and patient-reported data

IIVANAINEN, S.[1] EKSTRÖM, J.[2] VIRTANEN, H.[2] KATAJA, V.[2] LANG, L.[2] AND KOIVUNEN, J.[1]

1. Department of Oncology and Radiotherapy, Oulu University Hospital (OYS), MRC Oulu, Oulu, Finland
2. Kaiku Health, Helsinki, Finland

Conflict of interest: SI declares no conflict of interest; JK is an advisor for Kaiku Health Oy

BACKGROUND

- Immune checkpoint inhibitors (ICIs) are a standard of care treatments in several malignancies both in adjuvant and advanced settings. However, the treatment response assessment of ICIs differs from traditional cancer therapies with unique tumor response patterns such as pseudo- and hyperprogression. Furthermore, the temporal association of radiological response to treatment may sometimes be obscure[1].
- While only a subset of patients respond to ICIs, novel tools to assess the treatment response are needed aiming to improve patient-care and clinical value of ICIs.
- The prognostic role of immune-related adverse events (irAEs) implies that a niche of patients who benefit from ICIs can be identified[2,3]. A comprehensive and timely assessment of patients´ symptoms undergoing ICI therapies is feasible via electronic (e) patient-reported outcomes (PROs) collection[4]. We have previously shown that ePRO data can be combined with other clinical data sources to generate machine learning (ML) based models which predict irAEs[5, 6].
- The aim of this study was to investigate whether it is possible to predict objective response rate (ORR) in patients undergoing ICIs for advanced cancers using clinical and ePRO data as an input for a ML model.

METHODS

ORR was defined as the proportion of patients in whom partial (PR) or complete (CR) responses were seen as the best overall response (BOR) according to Response Evaluation Criteria in Solid Tumors (RECIST 1.1). Stable disease (SD) was categorized as non-response together with progessive disease (PD). ML-based prediction model for ORR prediction was built by using data collected from 31 patients with advanced cancers receiving ICI therapies in Oulu University Hospital. Several data sources were used as inputs for the model:

- Clinician-assessed treatment responses (n=63) according to the RECIST 1.1
- Clinician confirmed immune-related adverse events (irAEs) according to Common Terminology Criteria for Adverse Events (CTCAE) v.5.0
- Patient-reported symptom data including 18 monitored symptoms collected using the Kaiku Health digital platform
- Laboratory measurements from 9 different tests including bilirubin, hemoglobin, ALP, ALT, platelets, leukocytes, creatinine, thyrotropin and neutrophils
- Other variables: time from treatment initiation, age and sex

Treatment responses and irAEs were collected prospectively. Closest preceding lab values and reported symptoms, both as changes from the baseline, were linked to the treatment responses. In addition, the model accounted whether the patient had had a diagnosed irAE prior/at the time of response evaluation.

The prediction model for ORR was built using extreme gradient boosting (XGBoost algorithm), which is a commonly used approach for classification problems[7]. The complete modeling framework is explained in Figure 1. Treatment responses according to RECIST 1.1 were divided into binary categories, ie. objective response (PR+CR) vs no objective response (SD+PD). The binary categories predicted were the following:

- Complete response (CR) or partial response (PR), i.e. patient has an objective response
- Stable disease (SD) or progressive disease (PD), i.e. patient does not have an objective response

Prediction performance of the model for unseen samples was evaluated using leave-one-out cross-validation (LOOCV), which trained and tested 63 models, each time iteratively leaving one sample out as a test set. The LOOCV prediction performance was evaluated with accuracy, AUC (Area Under Curve), F1 score and MCC (Matthew's correlation coefficient). The performance metrics are presented in detail in Table 1.

TABLE 1. *Metrics used to evaluate the performance of the ORR prediction model.*

Metric	Description	Values
Accuracy	Describes how many predictions were correct as percentages.	0–100%. 100% indicates perfect classification.
AUC	Describes how well a model can distinguish between two classes (objective response OR non-response). Common performance metric for binary classification.	Gets values between 0 and 1. 1 is perfect classification and 0.5 is random guessing
F1 score	Harmonic mean of two commonly used metrics, precision and recall[8].	Gets values between 0 and 1. 1 indicates a perfect precision and recall.
MCC	Summarizes all possible cases for binary predictions: true and false positives and true and false negatives. Suitable for analyzing imbalanced datasets, where one class is rarer than the other.[9]	Gets values between -1 and 1. 1 is a perfect classification, 0 is random guessing and -1 indicates a completely contradictory classification.

RESULTS

The ORR prediction model had a promising LOOCV performance with all four metrics. The assessed metrics are presented in Table 2. Figure 2 presents a confusion matrix combining all 63 LOOCV predictions and Figure 3 illustrates the feature importances from a model trained with all available samples.

TABLE 2. *XGBoost LOOCV performance metrics for predicting ORR.*

	Accuracy	AUC	F1 Score	MCC
ORR prediction performance using LOOCV	75%	0,71	0,58	0,4

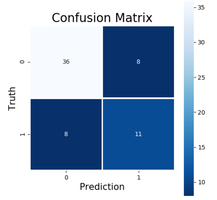

FIGURE 2. *Confusion matrix for predicted ORR. Upper left corner shows correctly classified negative, lower right corner correctly classified positive, upper right corner false positive and lower left corner false negative samples. Negative samples consist of SD and PD responses and positive samples CR and PR responses.*

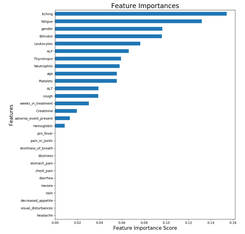

FIGURE 3. *Feature importances of ORR prediction model trained with all available samples. The displayed importances depict the relative average improvement in prediction accuracy across all 100 trees in the model where a certain feature is utilized. The importances should be considered as relative to each other.*

CONCLUSIONS

These promising results indicate that ML models built using ePRO symptom data and other clinical data can be used in treatment response prediction even with a limited size cohort.

AI models of the study offer a change for individually evolving follow-up with a possibility for prediction of important clinical events such as therapy toxicities and/or benefits.

REFERENCES

1. E. Borcoman, Y. Kanjanapan, S. Champiat, S. Kato, V. Servois, R. Kurzrock, S. Goel, P. Bedard and C. Le Tourneau: Novel patterns of response under immunotherapy, Ann Oncol 30, 385-396 (2019). 10.1093/annonc/mdz003 [doi]
2. C. Morehouse, S.E. Abdullah, M Dar, K Ranade & B. Higgs: Early incidence of immune-related adverse events (irAEs) predicts efficacy in patients (pts) with solid tumors treated with immune-checkpoint inhibitors (ICIs), J Clin Oncol 37(15):S2653 (2019). 10.1200/JCO.2019.37.15_suppl.2563 [doi]
3. H.T. Quach, A.K. Dewan, E.J. Davis, K.A. Ancell, R. Fan & F. Ye, et al.: Association of Anti–Programmed Cell Death 1 cutaneous toxic effects with outcomes in patients with advanced melanoma, JAMA Oncol 5(6):906–8 (2019).
4. S. Iivanainen, T. Alanko, P. Vihinen, T. Konkola, J Ekström, H Virtanen & J. Koivunen: Follow-Up of Cancer Patients Receiving Anti-PD-(L)1 Therapy Using an Electronic Patient-Reported Outcomes Tool (KISS): Prospective Feasibility Cohort Study. JFR 4(10):e17898 (2020). 10.2196/17898 [doi]
5. S. Iivanainen, J. Ekström, H. Virtanen, V. Kataja and J. Koivunen: Predicting the onset of immune-related adverse events (irAEs) in immune checkpoint inhibitor (ICI) therapies using a machine learning (ML) model trained with electronic patient-reported outcomes (ePROs) and lab measurements, Ann Oncol 31 (suppl_4): S1057 (2020). 10.1016/j.annonc.2020.08.1488 [doi]
6. S. Iivanainen, J. Ekström, V. Kataja, H. Virtanen and J. Koivunen: Electronic patient-reported outcomes (ePROs) and machine learning (ML) in predicting the presence and onset of immune-related adverse events (irAEs) of immune checkpoint inhibitor (ICI) therapies, J Clin Oncol 38, e14058-e14058 (2020). 10.1200/JCO.2020.38.15_suppl.e14058 [doi]
7. T. Chen and C. Guestrin: XGBoost: A Scalable Tree Boosting System: In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). ACM, New York, NY, USA, 785-794 (2016). 10.1145/2939672.2939785 [doi]
8. Z. Lipton, C. Elkan and B. Naryanaswamy: Optimal Thresholding of Classifiers to Maximize F1 Measure, Mach Learn Knowl Discov Databases 8725, 225-239 (2014). doi:10.1007/978-3-662-44851-9_15 [doi]
9. D. Chicco: Ten Quick Tips for Machine Learning in Computational Biology: BioData Min. 10:35 (2017). 10.1186/s13040-017-0155-3 [doi]

IN COLLABORATION

CONTACT DETAILS

sanna.iivanainen@ppshp.fi, tel. +358 8 315 3038
jussi.koivunen@ppshp.fi, tel. +358 8 315 3789

FIGURE 1:

A complete modeling framework behind the ORR prediction model.