Case Report

Predicting Onset and Continuity of Patient-Reported Symptoms in Patients Receiving Immune Checkpoint Inhibitor (ICI) Therapies Using Machine Learning

Sanna Iivanainen[1], Jussi Ekström[2], Henri Virtanen[2], Jussi Koivunen[1]*

[1]Department of Oncology and Radiotherapy, Oulu University Hospital and MRC Oulu, Oulu, Finland
[2]Kaiku Health Oy, Helsinki, Finland

***Corresponding Author:** Dr. Jussi Koivunen, Department of Oncology and Radiotherapy, Oulu University Hospital and MRC Oulu, Oulu, Finland, Tel: +35883153789; E-mail: jussi.koivunen@ppshp.fi

1. Introduction

In recent years, immune checkpoint inhibitor (ICI) therapies have become standard-of-care treatments in several advanced malignancies as well as adjuvant settings [1-10]. While ICIs provide new ways to treat a variety of cancer types, they also introduced novel toxicities. These toxicities can arise from various organ systems, and at any time point during treatments or even after treatment discontinuation [11-13]. The toxicities can be life threatening, but most are reversible if detected and treated early. Immune-related adverse events may also persist or appear in a similar manner after ICI discontinuation while immune-mediated toxicity seems to be independent of dose and duration of the given anti-PD-(L)1 treatment [14-16]. Due to the somewhat unpredictable nature of immune-related adverse events (irAEs), early detection of the toxicities is crucial and could result both in improved safety profile of treatments and a better quality-of-life for the patients.

Artificial intelligence (AI) based analytics have gained growing interest in the field of cancer care. Deep learning systems have shown promising results especially in cancer diagnostics [17]. AI based methods can be used to analyze vast data pools to create predictive analytics for generating value-based healthcare assets. Patient reported outcomes (PROs) consist of health-related questionnaires filled by the patients themselves which can capture symptoms and signs and their severity. Scheduled electronic (e) PROs have many advantages compared to paper questionnaires such as reducing timely and locational limitations and offering continuous collection of symptoms in a cost-effective manner [18-20]. Furthermore, ePROs could be used in the development of machine learning (ML) based approaches, such as symptom prediction models to enable earlier detection of toxicities.

The Kaiku Health digital platform has been used in a real-world setting to capture symptom data from cancer patients receiving ICIs. In this study, anonymized and aggregated ePRO data collected with Kaiku Health ePRO tool were used to train and tune models built using gradient boosting algorithm from an open source Python library XGBoost [21-23] for symptom continuity and onset prediction for 14 symptoms related to ICI toxicities.

2. Methods

The modelling methodology used in this study follows a general framework of data classification in machine learning. The outcome to be predicted is binary, i.e., symptom will onset or continue in the upcoming days or symptom won't onset or continue in the upcoming days. The used dataset consisted of ePRO data from ICI-treated patients collected using the Kaiku Health platform. There were 18 monitored symptoms in the ePRO questionnaire. The original dataset, which was used in the prediction model training, tuning and validation consisted of 21 744 reported symptoms from 72 ICI patients. This dataset was split into two using 70% of the data in training and tuning and 30% of the data in the initial validation of the models. The test dataset for the model performance evaluation in this study contained 16 884 reported symptoms collected with Kaiku ePRO tool from 67 cancer patients receiving ICIs which was collected after the model training to evaluate the performance of the models in real-life settings.

The symptom prediction models were built using an open source Python library XGBoost, which is an ensemble of many, usually hundreds of classification and regression trees known as CARTs. In this study, the tree booster of XGBoost library was used to train models for classification task of detecting symptoms related to ICIs. The model features were extracted from the past values of the 18 monitored symptoms (the predicted symptom and 17 other symptoms) to capture the variations in symptoms. The most suitable number of past values was found out to be three. In addition, differences in the symptom grades between the two previous values and the time between the reported data samples (for the three previous values) were included in the features.

XGBoost has several tunable parameters called hyperparameters which can be optimized to improve the performance of the trained models. In this study, hyperparameter tuning was done by using grid search with repeated (five repeats) stratified 5-fold cross-validation and logarithmic loss as the scoring metric in cross-validation. Scaling ratio for positive samples was calculated based on the size of the positive class compared with the negative class, and thus, it was not tuned with other parameters in the hyperparameter tuning. Also, learning rate of the ensemble was fixed to 0.01, and thus, it was not tuned with the other parameters.

The performance of the prediction was evaluated based on four known metrics. Accuracy describes how many predictions were correct as percentages, and 100% indicates perfect classification. Area under curve (AUC) is a performance metric for binary classification ranging from 0 to 1. F1-score is the weighted average of precision (how many of the cases predicted as positive are actually positive) and recall (how many of the positive cases are detected) which gets values between 0 and 1. Matthew's correlation coefficient (MCC) summarizes all possible cases for binary predictions: true and false positives, and true and false negatives. MCC is also suitable for analyzing

imbalanced datasets, where other class is much rarer than the other. MCC can be considered as a correlation coefficient between observed and predicted classifications and it gets values between -1 and 1, where 1 is perfect classification, 0 is random guessing and -1 indicates a completely contradictory classification. The complete modelling framework is presented in Figure 1.

Figure 1: Flowchart of the complete modelling framework.

3. Results

There were 18 monitored symptoms in the ePRO questionnaire, but four of them had too few reported cases in the dataset for prediction model training. Therefore, separate XGBoost models were trained for 14 ICI related

symptoms. Performance metrics for the 14 prediction models are presented in Table 1. The overall performance of the models was good, and the best performance considering all metrics was found for dyspnea, joint pain, cough and fatigue whereas the overall performance was lowest for nausea, headache, diarrhea and fever.

According to AUC values, the best performing models were for dyspnea and joint pain. The lowest AUC values were for headache and diarrhea. Generally, all the models were performing at a good level and the average AUC over the 14 models was 0,86. According to the accuracy score, all models were performing at a very good level; however, some symptoms had only few positive samples (cases where symptoms were present), which can distort the results. Such symptoms include, e.g., fever (only 3% of positive samples) and diarrhea (7%). F1-scores indicate that the best performing models were dyspnea and fatigue. The worst performance was found for fever and diarrhea. For these rarer symptoms, the lower F1-scores were caused by the higher number of false positives. According to MCC, the best performing models were for dyspnea and joint pain. The worst performance was found for nausea and headache. The overall performance was on a good level and the average MCC over the 14 models was 0,54 (Table 1).

Predicted symptom	AUC	Accuracy [%]	F1-score	MCC	Percentage of positive samples [%]
Dizziness	0,88	87,63	0,66	0,59	17
Itching	0,84	88,71	0,64	0,58	15
Fever	0,9	97,72	0,41	0,4	3
Diarrhea	0,78	89,78	0,44	0,4	7
Stomach pain	0,86	92,47	0,63	0,59	10
Nausea	0,82	83,2	0,45	0,36	14
Fatigue	0,9	82,12	0,83	0,64	50
Rash	0,84	88,58	0,55	0,49	11
Decreased appetite	0,84	84,81	0,54	0,47	13
Cough	0,9	86,56	0,78	0,69	30
Dyspnea	0,96	91,4	0,85	0,79	28
Joint pain	0,92	89,11	0,82	0,75	29
Headache	0,77	82,93	0,49	0,39	16
Chest pain	0,88	88,71	0,51	0,46	9

Table 1: Performance metrics for the 14 prediction models.

4. Discussion

Artificial intelligence (AI) or machine intelligence is intelligence demonstrated by machines. In health care, knowledge representation as part of the clinical decision support system is currently the most used AI approach. There are high hopes that AI could improve health care with early diagnostics and improved care in a more cost-

effective manner compared to current approaches. The intriguing aspect of AI is its' capability to analyze vast amounts of data, and based on that, detect correlations. The possibility of utilizing AI in the ePRO follow-up of treatment toxicities in that it could detect developing severe symptom cascades, and thus, instruct physicians and patients in advance, is extremely interesting. Whether this could be enhanced by combining the symptom reports of a patient to other eHealth apps sensing for example metabolic or physiologic changes, is another fascinating possibility [24].

This study found that ML based modeling of ePRO data on ICI treated cancer patients is feasible in predicting the onset and continuation of symptoms related to ICI toxicities. Advisable patient education and communication can improve quality of cancer care with multiple subjective and objective enhancements [25, 26]. In clinical practice, applications for the prediction of the continuity and onset of symptoms could provide proactive support for patients through, e.g., timely patient education and guidance. Furthermore, ePRO tools coupled with AI analytics could enable more precise follow-up by timing symptom questionnaires based on predictions, thus, a more personalized, and most likely cost-effective, risk-based follow-up.

Toxic effects of conventional chemotherapy and molecularly targeted cancer therapies are generally well defined and occur at predictable points. By contrast, due to somewhat unpredictable timing, and clinical overlap with other conditions, immune-related adverse events (irAE) may be more difficult to diagnose and characterize [27-29]. The study also suggests that ML based prediction models could be utilized in the early detection of ICI related toxicities using ePRO data of symptoms related to ICI toxicities. There is growing evidence that patients treated with immune checkpoint inhibitors developing irAE are more likely to benefit from the therapy [30-34]. According to a study on lung cancer follow-up, ePROs enable cost-effective capturing of symptoms and their change over long time-periods [20]. Changes over time might better predict treatment side effects and benefit than just a single presentation of a symptom. Large scale symptom data combined with treatment benefit and side effects could be used to build prediction models using ML based methods. The models could predict risk for an individual patient for symptom development, treatment related side effects, and treatment benefit.

This study revealed that ML based predictive analytics on the onset and presence of ICI related symptoms based on ePRO data is feasible. The results of the study are encouraging and suggest that ML based prediction models on ePROs of irAE data could be further utilized in the early detection of ICI related toxicities. In the future, the prediction models should be validated with a dataset collected in a prospective clinical trial to assess, if they truly are adding value to cancer care.

Disclosure of Potential Conflicts of Interest

Ethics Approval and Consent to Participate

According to national legislation, informed consent is not needed due to the retrospective and non-interventional nature of the study.

Availability of Data and Materials

The datasets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Authors' Contributions

SI and JPK contributed to the conception and design of the study; SI, JPK, HV, and JE acquired the data; JE analyzed the data and JE, SI and JPK interpreted the data. SI, JE and JPK contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Funding

References

1. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial, Lancet 387 (2016): 1540-1550.

2. Brahmer J, Reckamp KL, Baas PL, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer, N Engl J Med 373 (2015): 123-135.

3. Robert C, Long GV, Brady B, et al. Nivolumab in previously untreated melanoma without BRAF mutation, N Engl J Med 372 (2015): 320-330.

4. Motzer RJ, Escudier B, McDermott DF, et al. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma, N Engl J Med 373 (2015): 1803-1813.

5. Motzer RJ, Tannir NM, McDermott DF, et al. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma, N Engl J Med 378 (2018): 1277-1290.

6. Bellmunt J, de Wit R, Vaughn DJ, et al. Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma, N Engl J Med 376 (2017): 1015-1026.

7. Robert C, Schachter J, Long GV, et al. Pembrolizumab versus Ipilimumab in Advanced Melanoma, N Engl J Med 372 (2015): 2521-2532.

8. Rittmeyer F, Barlesi D, Waterkamp K, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial, Lancet 389 (2017): 255-265.

9. Weber J, Mandala M, Del Vecchio M, et al. Adjuvant Nivolumab versus Ipilimumab in Resected Stage III or IV Melanoma, N Engl J Med 377 (2017): 1824-1835.

10. Eggermont MM, Blank CU, Mandala M, et al. Adjuvant Pembrolizumab versus Placebo in Resected Stage III Melanoma, N Engl J Med 378 (2018): 1789-1801.

11. Haanen JBAG, Carbonnel F, Robert C, et al. Management of toxicities from immunotherapy: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up, Ann Oncol 28 (2017): iv119-iv142.

12. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer, N Engl J Med 366 (2012): 2443-2454.

13. Wang DY, Salem JE, Cohen JV, et al. Fatal Toxic Effects Associated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis, JAMA Oncol 4 (2018): 1721-1728.

14. Spain L, Diem S, Larkin J. Management of toxicities of immune checkpoint inhibitors, Cancer Treat Rev 44 (2016): 51-60.

15. Puzanov A, Diab K, Abdallah CO, et al. Managing toxicities associated with immune checkpoint inhibitors: consensus recommendations from the Society for Immunotherapy of Cancer (SITC) Toxicity Management Working Group, J Immunother Cancer 5 (2017): 95-z.

16. Brahmer R, Lacchetti C, Schneider BJ, et al. Management of Immune-Related Adverse Events in Patients Treated With Immune Checkpoint Inhibitor Therapy: American Society of Clinical Oncology Clinical Practice Guideline, J Clin Oncol 36 (2018): 1714-1768.

17. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening, Nature 577 (2020): 89-94.

18. Holch P, Warrington L, Bamforth LCA, et al. Development of an integrated electronic platform for patient self-report and management of adverse events during cancer treatment, Ann Oncol 28 (2017): 2305-2311.

19. Kotronoulas G, Kearney N, Maguire R, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials, J Clin Oncol 32 (2014): 1480-1501.

20. Lizee T, Basch E, Tremolieres P, et al. Cost-Effectiveness of Web-Based Patient-Reported Outcome Surveillance in Patients With Lung Cancer, J Thorac Oncol 14 (2019): 1012-1020.

21. Chen T, Guestrin C. 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 (2016): 785-794.

22. Friedman H. Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics 29 (2001): 1189-1232.

23. XGBoost. Scalable, Portable and Distributed Gradient Boosting (GBDT, GBRT or GBM) Library.

24. Abbasi J. "Electronic Nose" Predicts Immunotherapy Response, JAMA 322 (2019): 1756.

25. LeBlanc TW, Abernethy AP. Patient-reported outcomes in cancer care - hearing the patient voice at greater volume, Nat Rev Clin Oncol 14 (2017): 763-772.

26. Gilligan T, Coyle N, Frankel RM, et al. Patient-Clinician Communication: American Society of Clinical Oncology Consensus Guideline, J Clin Oncol 35 (2017): 3618-3632.

27. Hsiehchen D, Watters MK, Lu R, et al. Variation in the Assessment of Immune-Related Adverse Event Occurrence, Grade, and Timing in Patients Receiving Immune Checkpoint Inhibitors, JAMA Netw Open 2 (2019): e1911519.

28. Fares CM, Williamson TJ, Theisen MK, et al. Low Concordance of Patient-Reported Outcomes with Clinical and Clinical Trial Documentation, JCO Clin Cancer Inform 2 (2018): 1-12.

29. So C, Board RE. Real-world experience with pembrolizumab toxicities in advanced melanoma patients: a single-center experience in the UK, Melanoma Manag 5, MMT05-0028. eCollection (2018).

30. Fujii T, Colen RR, Bilen MA, et al. Incidence of immune-related adverse events and its association with treatment outcomes: the MD Anderson Cancer Center experience, Invest New Drugs 36 (2018): 638-646.

31. Liew DFL, Leung JLY, Liu B, et al. Association of good oncological response to therapy with the development of rheumatic immune-related adverse events following PD-1 inhibitor therapy, Int J Rheum Dis 22 (2019): 297-302.

32. Martini DJ, Hamieh L, McKay RR, et al. Durable Clinical Benefit in Metastatic Renal Cell Carcinoma Patients Who Discontinue PD-1/PD-L1 Therapy for Immune-Related Adverse Events, Cancer Immunol Res 6 (2018): 402-408.

33. Sznol MA, Postow MJ, Davies AC, et al. Endocrine-related adverse events associated with immune checkpoint blockade and expert insights on their management, Cancer Treat Rev 58 (2017): 70-76.

34. Haratani K, Hayashi H, Chiba Y, et al. Association of Immune-Related Adverse Events With Nivolumab Efficacy in Non-Small-Cell Lung Cancer, JAMA Oncol 4 (2018): 374-378.