Predicting symptom trajectories among ambulatory cancer patients receiving anticancer treatment using machine learning approaches: A feasibility study

UCI School of Pharmacy & Pharmaceutical Sciences
UCI Donald Bren School of Information & Computer Sciences
UCI Chao Family Comprehensive Cancer Center

Ding Quan Ng[1], Yawen Guo[2], Rukh Yusuf [1], Daniela Arcos[1], Alison Chen[3], Benjamin Lee[1,3], Lan Duong[3], Linda Van[3], Thomas Nguyen[3], Vuong Green[3], Daniel Hoang[1,3], Kai Zheng[2], Alexandre Chan[1,3]

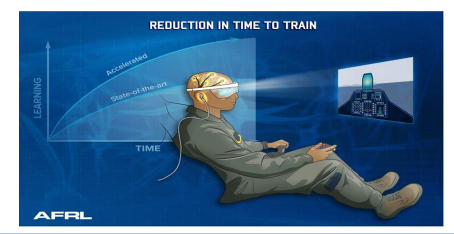
[1] School of Pharmacy & Pharmaceutical Sciences, University of California Irvine; [2] Donald Bren School of Information and Computer Sciences, University of California Irvine;
[3] Department of Pharmacy, Chao Family Comprehensive Cancer Center.

Background

Machine learning (ML) approach can predict symptom trajectories which could allow clinicians to prevent worsening symptoms at point-of-care using preventive interventions and personalized symptom management advice.

Objective

- Examine the feasibility of developing ML models to predict cancer-related symptoms of fatigue, pain interference, anxiety, depression, nausea and vomiting, as well as physical and cognitive function.
- Implement ML into a clinician-accessible dashboard for displaying symptom severity and predicted future events.



Methods

- Diagnosed cancer patients completed a series of NIH **PROMIS® questionnaires** at the Chao Family Comprehensive Cancer Center (UCI IRB #20216431) on various cancer-related symptoms including fatigue, pain interference, anxiety, depression, nausea and vomiting, as well as physical and cognitive function.
- **PROMIS symptom scores** are generated with each questionnaires **representing the severity of symptoms**
- If a patient used the toolkit on multiple visits, we rely on **prior visit data** to predict the PROMIS symptom scores of **the subsequent visit**.
- The ML models were fed with features from the prior visit, including:

Baseline survey data	PROMIS scores from the previous visit
Clinical characteristics from the first visit	Routine clinical biomarkers obtained via the UC Health Data Warehouse using the UCI Health Honest Broker service

- **Linear Regression, Support, Vector Regression, Random Forest Regression, and Gradient Boosting Regression** were employed on processed data.
- Please refer to Figure 1 for the diagrammatic representation of the study design.

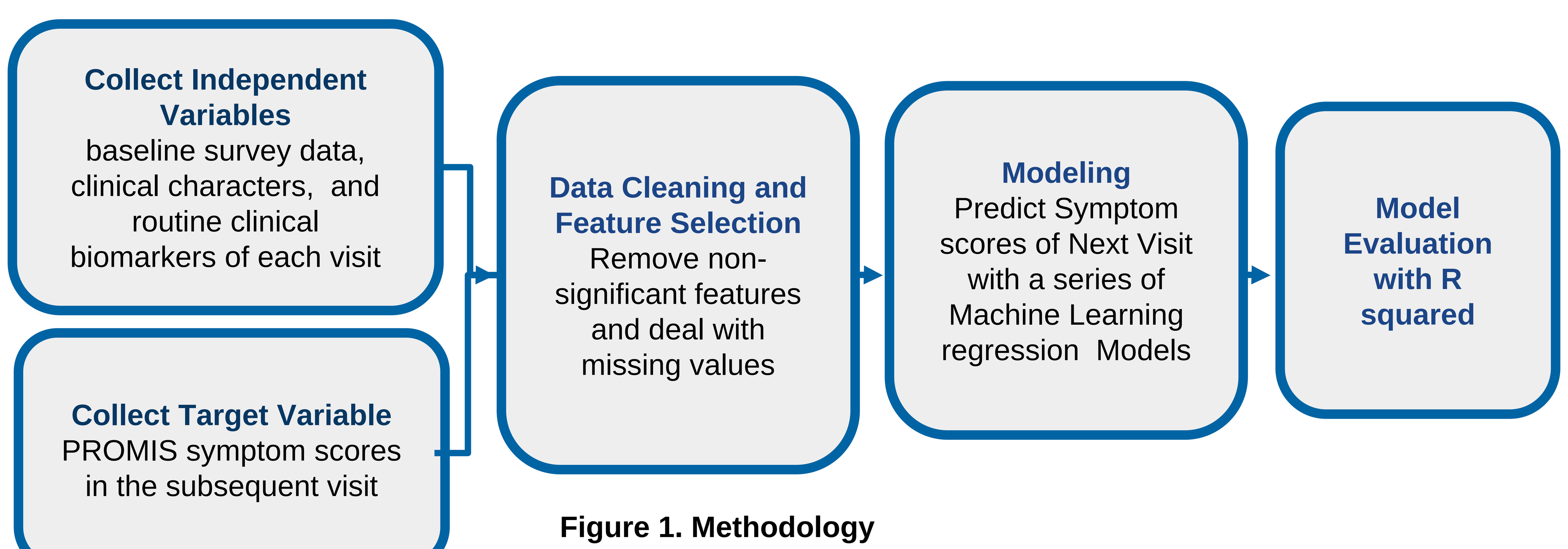


Figure 1. Methodology

Results

- A total of 289 patient visit record, of which 144 contained data of a subsequent visit necessary for model development.
- Model performance evaluated using R^2 values is reported in Table 1. **Linear regression consistently outperformed other models for all target symptoms.**
- Table 2 describes the top five significant features of the linear regression model necessary for accurate predictions. **They vary depending on the symptom assessed and comprise immuno-oncology agents, clinical biomarkers and sociodemographic characteristics.**

Table 1. R^2 on Prediction Models and Dependent Symptom Variables

R^2 Performance	Linear Regression	SVR(linear kernel)	Random Forest Regression	Gradient Boosting Regression
Anxiety	**0.55**	0.45	0.07	-0.01
Depression	**0.57**	0.48	0.15	0.14
Cognitive Function	**0.46**	0.35	0.23	0.05
Physical Function	**0.54**	0.47	0.20	-0.04
Fatigue	**0.48**	0.43	0.17	0.09
Pain Interference	**0.39**	0.28	0.20	0.02
Nausea and Vomiting	**0.42**	0.30	0.23	0.12

Abbreviation(s): SVR, support vector regression.

Table 2. Top Significant Features of the Best Performing Linear Regression Model

R^2 Performance	Top five features (in order, from left to right)
Anxiety	***Nivolumab***, red blood cell count, albumin, creatinine, exposure to immunotherapy
Depression	***Durvalumab***, creatinine, red blood cell count, exposure to immunotherapy, race & ethnicity
Cognitive Function	***Nivolumab***, pembrolizumab, sex, cemipilimab, atezolizumab
Physical Function	***Cemipilimab***, atezolizumab, ipilimumab, red blood cell count, creatinine
Fatigue	***Creatinine***, race & ethnicity, marital status, cemipilimab, atezolizumab
Pain Interference	***Red blood cell count***, nivolumab, creatinine, sex, marital status
Nausea and Vomiting	***Nivolumab***, creatinine, red blood cell count, sex, albumin

Conclusions

- The feasibility of utilizing ML to predict medical symptoms has been demonstrated in our study.
- Various models may perform differently with some models more effective than others, there is no one-size-fits-all solution.
- Various features are better at predicting some but not other symptoms.

Future Directions

- We will continue to refine current models through parameter tuning and independent feature selection targeting different symptoms.
- The set of clinically relevant features will also expand to incorporate cancer diagnoses and anticancer drugs.
- Finally, we will explore the prediction of other health outcomes such as unplanned healthcare utilizations.

References

- Bevans M, et al. Nursing Outlook. 2014;62(5):339-345.
- Glasgow RE, et al. Am J Public Health. 1999 Sep;89(9):1322-7.

Acknowledgement

This work is supported by the **Hematology/Oncology Pharmacy Association (PI: Alexandre Chan)**