Accepted Manuscript

Machine Learning Methods to Extract Documentation of Breast Cancer Symptoms from Electronic Health Records

Alexander W. Forsyth, MEng, Regina Barzilay, PhD, Kevin S. Hughes, MD, Dickson Lui, MB ChB, Karl A. Lorenz, MD, Andrea Enzinger, MD, James A. Tulsky, MD, Charlotta Lindvall, MD, PhD

PII: S0885-3924(18)30082-4

DOI: 10.1016/j.jpainsymman.2018.02.016

Reference: JPS 9733

To appear in: *Journal of Pain and Symptom Management*

Received Date: 16 November 2017

Revised Date: 15 February 2018

Accepted Date: 19 February 2018

Please cite this article as: Forsyth AW, Barzilay R, Hughes KS, Lui D, Lorenz KA, Enzinger A, Tulsky JA, Lindvall C, Machine Learning Methods to Extract Documentation of Breast Cancer Symptoms from Electronic Health Records, *Journal of Pain and Symptom Management* (2018), doi: 10.1016/j.jpainsymman.2018.02.016.

Machine Learning Methods to Extract Documentation of Breast Cancer Symptoms from Electronic Health Records

Alexander W. Forsyth, MEng[1]; Regina Barzilay, PhD[1]; Kevin S. Hughes, MD[2]; Dickson Lui, MB ChB[3]; Karl A. Lorenz, MD[4,5]; Andrea Enzinger, MD[6,7,8]; James A. Tulsky, MD[7,8]; Charlotta Lindvall, MD, PhD[7,8]

[1]Department of Electrical Engineering and Computer Science, CSAIL, MIT, Cambridge, USA; [2]Department of Surgical Oncology, Massachusetts General Hospital; [3]Department of Medicine, Waitemata District Health Board, Auckland, New Zealand; [4]Primary Care and Population Health, Department of Medicine, Stanford School of Medicine, Stanford, USA; [5]VA Palo Alto Health Care System, Palo Alto, USA; [6]Division of Population Sciences, Dana-Farber Cancer Institute, Boston, USA; [7]Department of Psychosocial Oncology and Palliative Care, Dana-Farber Cancer Institute, Boston, USA; [8]Division of Palliative Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, USA

Address for correspondence:
Charlotta Lindvall, MD, PhD
Department of Psychosocial Oncology and Palliative Care
Dana-Farber Cancer Institute
450 Brookline Ave, LW670
Boston, MA 02215-5450
Phone: 617-582-9317
Fax: 617-632-6180
Email: Charlotta_lindvall@DFCI.harvard.edu

Running title: Machine learning based symptom identification

Word count: 2,983

Key words: Machine learning; natural language processing; patient-reported symptoms; electronic health record; breast cancer; palliative care