

**Title** (161/255 characters)

## **Evaluating Performance of an Artificial Intelligence Tool to Identify Individuals in Need of Three-Year Colorectal Cancer Post-Polypectomy Surveillance**

Jamie Yang<sup>1</sup>, Sadie DeSilva<sup>2</sup>, Brandon Smith<sup>2</sup>, Camille Soroudi<sup>2</sup>, Anthony Myint<sup>2</sup>, Bitu Shahrivini<sup>1</sup>, William Hsu<sup>3</sup>, Cleo Maehara<sup>3</sup>, Jayraan Badiie<sup>2</sup>, Alexandria Uy<sup>3</sup>, Yuna Kang<sup>4</sup>, Bitu Naini<sup>4</sup>, Venkataraman Muthusamy<sup>2</sup>, Eric Esrailian<sup>2</sup>, Folasade P. May<sup>2,5,6</sup>

<sup>1</sup> UCLA Department of Internal Medicine, Los Angeles, CA, USA

<sup>2</sup> Vatche and Tamar Manoukian Division of Digestive Diseases; UCLA David Geffen School of Medicine

<sup>3</sup> UCLA Department of Radiological Sciences; Data Integration, Architecture and Analytics Group

<sup>4</sup> UCLA Department of Pathology and Laboratory Medicine

<sup>5</sup> Greater Los Angeles Veterans Affairs Health System, Los Angeles, CA, USA

<sup>6</sup> UCLA Kaiser Permanente Center for Health Equity, Jonsson Comprehensive Cancer Center

### **Category:**

AGA: Clinical Practice- Artificial Intelligence and/or Machine Learning

Limit: 2896/2900 characters

Deadline: 11/30 9:00 pm EST

**Introduction:** Post-polypectomy surveillance aims to increase early detection and prevention of colorectal cancer (CRC). High risk neoplasia (HRN), defined as adenoma or sessile serrated polyp (SSP) greater than 1 centimeter, adenoma with tubulovillous or villous histology, high-grade dysplasia, >5 adenomas or SSPs, or traditional serrated adenoma, increase CRC risk and require a 3-year surveillance interval as per the 2020 United States Multi-Society Task Force (MSTF) guidelines. We previously developed and implemented a natural language processing (NLP) algorithm that uses artificial intelligence (AI) to extract data from colonoscopy and pathology reports to determine guideline-concordant surveillance intervals. In this study, we aimed to evaluate the NLP's false negative rate for 3-year surveillance to help assess its overall performance before broad application in a health system intervention.

**Methods:** The setting for this study is a large academic health center with a defined primary care population, robust referral-based care, and 6 outpatient endoscopy facilities that perform over 15,000 screening colonoscopies annually. We randomly selected 1100 patients with screening colonoscopy exams performed between 1/1/2022 and 12/21/2022. We excluded patients with diagnostic colonoscopies. We compared the NLP-assigned post-polypectomy surveillance intervals to manual chart review with MSTF intervals as the gold standard. All discrepancies were reviewed by two physicians. We then calculated the false negative rate of the NLP model for the subgroup of patients requiring 3-year surveillance.

**Results:** Our overall study population consisted of 967 patients (50.4% female, 68.7% non-White, and mean age  $57.2 \pm 11.5$  years). There were 104 patients (10.7%) with a HRN diagnosis by manual chart review, compared to 121 patients (12.5%) by NLP. In total, 7 HRN patients (6.7%) were not detected by the NLP (false negative rate). In these 7 missed cases, 4 were misclassified by the NLP to longer surveillance intervals, 2 cases were misclassified to 1-year follow up, and 1 case could not be classified by NLP. Compared to manual chart review, the NLP was 93.3% sensitive and 97.2% specific for identifying patients indicated for 3-year surveillance.

**Conclusion:** We successfully developed an NLP algorithm in a large health system that uses AI to automatically identify patients with HRN for 3-year surveillance colonoscopy. The false negative rate was 6.7%. Given that timely colonoscopy surveillance is essential for early detection and removal of polyps, particularly for this high-risk group of individuals, this study highlights the potential of AI-based tools to promote guideline-based colonoscopy surveillance intervals and improve CRC outcomes. Future goals are to use this tool to automate recall and scheduling of HRN patients due for surveillance in our health system.

Table 1: Characteristics of Overall Study Population

<b>Demographics</b>	<b>n (%) or mean (SD)</b>
<b>Mean Age</b>	57.2 (11.5)
<b>Gender</b>	
Male	480 (49.6)
Female	487 (50.4)
<b>Race/Ethnicity</b>	
Non-Hispanic White or Caucasian	270 (27.9)
Non-Hispanic Black or African American	58 (6.0)
Hispanic	126 (13.0)
Non-Hispanic Asian	140 (14.4)
Non-Hispanic Middle Eastern or North African	34 (3.5)
Non-Hispanic Other*	248 (25.6)
Unknown or Declined to Answer or Do not Identify with Race	91 (9.41)
<b>Insurance</b>	
Commercial	548 (56.7)
HMO	135 (14.0)
Medi-Cal	36 (3.7)
Medicare	190 (19.6)
Other/Unknown	58 (6.0)

\*Non-Hispanic Other includes American Indian or Alaskan Native, Multiple Races, and other race categories.

Table 2: Confusion Matrix Comparing NLP Identification of patients requiring 3-year surveillance versus Manual Chart Review

		<b>Manual Chart Review MSTF interval</b>	
		Three-year surveillance Indicated (High Risk Neoplasia Patients)	Non-three-year Surveillance Patients
<b>NLP Algorithm</b>	Three-year surveillance Indicated (High Risk Neoplasia Patients)	97	24
	Non-three-year Surveillance Patients	7	839
<b>Test Parameters for 3-year surveillance interval</b>			
Sensitivity		93.3%	
Specificity		97.2%	
Positive Predictive Value		80.1%	
Negative Predictive Value		99.1%	
Missed HRN percentage		6.73%	