E0173 - VALIDATION OF A DEEP MACHINE LEARNING TOOL TO DETERMINE INTRA-PROCEDURAL SCREENING COLONOSCOPY QUALITY INDICATORS IN AN ACADEMIC HEALTH SYSTEM



David Geffen School of Medicine

Artin Galoosian, MD, MA¹, Jamie O. Yang, MD¹, Emma K. Peterson, MS², Cleo Maehara, MD³, Jayraan Badiee, MPH, MS^{2,3}, Camille Soroudi, MD², Anthony Myint, MD², Yuna Kang, MD³, Bita Naini, MD³, Sadie De Silva, MD², Raman Muthusamy, MD¹, Eric Esrailian, MD, MPH¹, William Hsu, PhD³, Folasade P. May MD, PhD MPhil^{1,2,4,5}

[1] Department of Medicine, David Geffen School of Medicine, UCLA; [2] The Vatche and Tamar Manoukian Division of Digestive Diseases; [3] UCLA Medical and Imaging Informatics, Data Integration, Architecture and Analysis Group; Department of Medicine, VA Greater Los Angeles Healthcare System; [5] Jonsson Comprehensive Cancer Center, Cancer Prevention Control Research, UCLA

BACKGROUND

- High quality screening colonoscopy is the hallmark of effective cole (CRC) prevention.
- Measuring colonoscopy quality indicators is crucial for quality impr
- However, accurately and routinely measuring colonoscopy quality providers and health systems remains a challenge.

OBJECTIVES

- To develop a natural language processing (NLP) algorithm that util learning tools to automatically and accurately characterize four preindicators for screening colonoscopies.
- To validate the performance of the NLP algorithm against manual

METHODS

Study setting:

 Large academic healthcare system that performs >15,100 screet yearly in 6 endoscopy centers.

Study Population:

• We randomly selected 600 screening colonoscopies performed b 2/2021.

Analyses:

- We trained and developed an NLP algorithm to measure colonos (IND), bowel preparation (BP), cecal intubation (CI), and success (SCI) (Figure 1).
- We compared NLP-derived quality metrics and manual chart revi chart review as gold standard).

RESULTS

- NLP sensitivity ranged from 99.3% to 100.0% and specificity range 100.0% for all four studied quality indicators.
- Because of conflicting documentation by endoscopist in the same misclassified 2 cases for colonoscopy indication and 1 case for bo
- NLP misclassified 12 cases for successful cecal intubation, which either the endoscopist not mentioning the word "cecum" or docum ileum" instead of "cecum."
- NLP had perfect performance (100%) for cecal intubation.

	Table 1: NLP Performance on four colonoscopy quality indicators				Figure 1: Scher
lorectal cancer	Documentation of colonoscopy indication (IND)				
		NLP _			
	Manual review	"Screening" detected	"Non-screening" detected	lotal	
rovement efforts.	"Screening" detected	314	0	314	
	"Non-screening" detected	2	284	286	
/ indicators for	Total	316	284	600	Free-text
tilizes machine especified quality	Test characteristics	Sensitivity	99.3%		colonoscopy
		Specificity	100%		reports
		PPV	100%		
		NPV	99.4%		
		F1 Score	0.996		
		Accuracy	99.7%		
	Documentation of bowel preparation (BP)				
chart review.		NLP		Tatal	
	Manual review	BP documented	BP not documented	Iotal	
	BP documented	599	0	599	
	BP not documented	1	9	1	
	Total	600	0	600	
		Sensitivity	100%		
ening colonoscopies	Test characteristics	Specificity	97.5%		
		PPV	99.8%		
		NPV	N/A		
		F1 Score	0.999		
		Accuracy	99.8%		
	Doc	umentation of cecal intuba	ation (CI)		Figure 1: The NLP
between 6/2020-	NI P			the report as structu	
	Manual review	CI documented	Cl not documented	Total	
	CI documented	599	0	599	
	Cl not documented	0	1	1	
scopy indication sful cecal intubation	Total	599	1	600	
		Sensitivity	100%	000	
	Test characteristics	Specificity	100%		
		PPV	100%		indicators.
view (two physician		NPV	100%		• The NILD electrit
		E1 Score	1		
			100%		and report color
	S	uccessful cecal intubation	(SCI)		. Eutura direction
		N	P		
	Manual review	SCI documented	SCI not documented	Total	
	SCI documented	437	3	440	
ged from 94.3% to	SCI not documented	407 Q	150	159	
	Total	116	153	599	UCLA
e report NI P		440	100	000	UCLA Va
aval proparation		Sensitivity	99.3%		
		Specificity	94.3%		
were mainly due to nenting "terminal	Test characteristics	PPV	98.0%		
		NPV	98.0%		Artin Galoosian, M Galoosian@gmail.co
		F1 Score	0.987		
		Accuracy	98.0%		

UCLA Health

matic of natural language processing pipeline



receives input from free-text colonoscopy reports and then outputs ured data that can be used to interpret quality indicators.

CONCLUSIONS

y developed an automated NLP algorithm that is highly accurate determining four priority intraprocedural colonoscopy quality

thm is a feasible, efficient, and effective method to consistently track noscopy quality metrics at the provider and health system level.

ns: application of our NLP algorithm to improve CRC outcomes.

ACKNOWLEDGEMENTS

A Melvin and Bren Simon Quality Improvement Program Atche and Tamar Manoukian Division of Digestive Diseases

CONTACT INFORMATION

ID, MA Folasade P. May MD, PhD, MPhil om FMay@mednet.ucla.edu

