Predicting Post-operative C. Difficile Infection (CDI) with Automated Machine Learning (AutoML) Algorithms Using the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) Database

Avinash Thangirala[1], Terry Li[1], Ehab Abaza[1] Yindalon Aphinyanaphongs[2], Jordan Axelrad[3], Ji Chen[2], Andrew C. Kelleher[1], Jacob F. Oeding[1], Eric Hu[1], Jacob A. Martin[4], Gregory Katz[4], Shelly Z. Brejt[4], Gabriel Castillo[3], Nicolai P. Ostberg[1], Karen Kan[4]

[1]Department of Medicine, NYU Langone Health, New York, NY, [2]Department of Population Health, NYU Langone Health, New York, NY, [3]Division of Gastroenterology, NYU Langone Health, New York, NY, [4]Division of Cardiology, NYU Langone Health, New York, NY

Introduction

- Clostridium difficile infection (CDI) is one of the most common hospital-acquired infections which leads to significant morbidity and increased healthcare costs.
- Surgical patients have an increased burden of CDI, resulting in poorer post-operative outcomes.
- Effective prediction can lead to preventative measures and/or closer monitoring for timely treatment
- Few prior studies have developed predictive models for CDI, most of which are logistic regression (LR) models.[1]
- Automated machine learning (AutoML) is an application of artificial intelligence which can build models with scale and efficiency and has consistently outperformed LR models in non-medical contexts.
- Autogluon, an AutoML system that has demonstrated superior benchmark results to other AutoML frameworks[2], has not yet been applied to predict 30-day mortality.

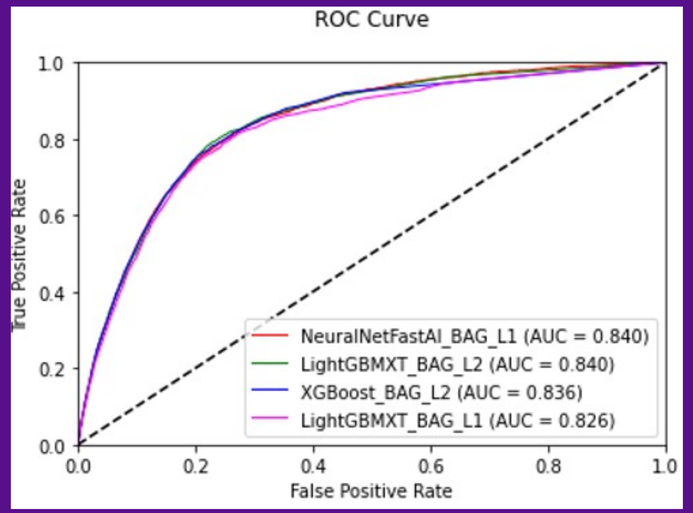
Aims

- **This study aims to investigate the utility of AutoML in developing a model for post-operative CDI risk stratification.**

Methods

- Utilized AutoML system by Amazon, Autogluon v0.3.1, and the 2016-2018 ACS NSQIP database
- All surgical types within the database
- 79 pre-operative features included in the model
- Post-operative CDI defined as CDI ≤ 30 days of surgery
- Models trained for 4 hours to optimize performance on Brier score, with lower being better
- Validation of all performance metrics done using the 2019 NSQIP database

Automated machine learning models offer similar or improved discriminatory characteristics to existing post-operative C. Difficile infection predictive models



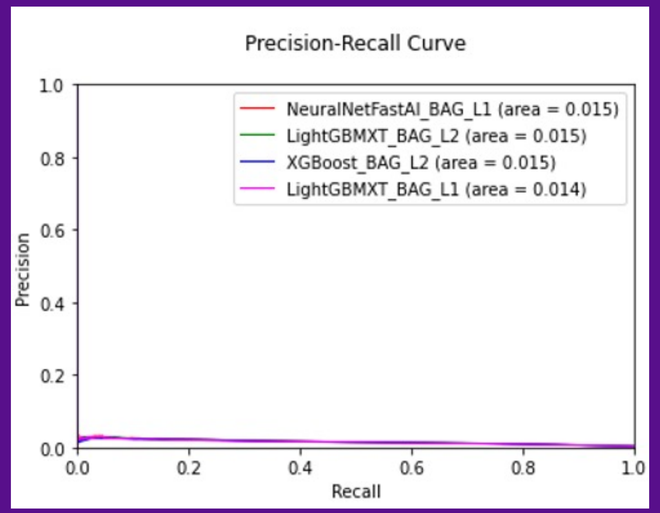


Figure 1: The Receiver Operating Curves (a) help visually evaluate the trade-off between true positive rate and false positive rate; whereas the Precision Recall Curves (b) help visually evaluate the trade-off between true positive rate and positive predictive value.

NYU Langone Health

NYU Langone Health

Results

Variable		Cases, n (%)	No	Yes	p-value
Sex	Female	1733472 (57)	1727523	5899	<0.001
	Male	1316142 (43)	1311040	5102	
Race	Indian/Alaskan	15673 (0.5)	15604	69	<0.001
	Asian	86454 (2.8)	86226	228	
	African/AA	307843 (10)	306727	1116	
	Hawaiian/PI	11148 (0.4)	11114	34	
	Unknown	498999 (16)	497574	1425	
	White	2129500 (70)	2316239	8129	
Surgical Specialty	Cardiac	12353 (0.4)	12278	75	<0.001
	General	1322545 (43)	1315721	6824	
	Gynecology	268647 (9)	268278	369	
	Neurosurgery	161615 (5)	161236	379	
	Orthopedics	729997 (24)	728691	1306	
	Urology	179142 (6)	178457	685	
	Vascular	163825 (5)	162864	961	
Elective Surgery	No	614917 (20)	609683	5234	<0.001
	Unknown	47066 (0.2)	4687	19	
	Yes	2429994 (80)	2424246	5748	
Smoker	No	2532971 (83)	2524114	8857	<0.001
	Yes	516646 (17)	514502	2144	
ASA Class	1	253948 (8)	253770	178	<0.001
	2	1356498 (45)	1354103	2395	
	3	1247668 (41)	1241603	6065	
	4	178613 (6)	176394	2219	
	5	5413 (0.2)	5288	125	
	None assigned	7477 (0.2)	7458	19	

Table 1: Baseline Characteristics of those with CDI (Yes) and without CDI (No)

- Top performing model was an ensembled neural net model and had a Brier score of 0.0027, an AUROC 0f 0.84, and an AUC-PR of 0.015 (Figure)

References

1. Woo SH, Hess B, Ackermann L, Cowan SW, Valentine J. Development and Validation of a web-based Postoperative Clostridioides difficile infection risk prediction model. Published online June 24, 2020:2020.06.23.20138420. doi:10.1101/2020.06.23.20138420
2. Hutter F, Caruana R, Bardenet R, Bilenko M, Guyon I, Kegl B, and Larochelle H. "AutoML 2014 @ ICML". *AutoML 2014 Workshop @ ICML*. Retrieved 2022-05-27.