Cary Cotton and Jared Sninsky

BIOS 635

Title: Clinical and Endoscopic Characteristics May Sufficiently Predict the Absence of Adenomas in Diminutive Rectosigmoid Polyps to Allow a Diagnose and Leave Strategy

Abstract

Although all polyps are routinely removed during colonoscopy, many diminutive rectosigmoid polyps have no malignant potential and increase healthcare costs for pathology examination and may increase sedation time substantially. Examination under narrow band imaging (NBI), a blue/green wavelength light available on modern endoscopes, has been shown to be a useful tool in expert hands to differentiate hyperplastic polyps from adenomas. Endoscopists untrained in NBI have not met the 90% negative predictive threshold goal published by the American Society of Gastrointestinal Endoscopy. We used demographic data, time from last colonoscopy, and concurrent gross colonoscopy findings to predict the presence of any diminutive (<5mm) rectosigmoid adenoma. We tuned a variety of machine learning models using repeated cross-validation in a training dataset: logistic regression with least absolute shrinkage and selection operator (LASSO), naive bayes, K-nearest neighbors (KNN), and random forest. Based on training set performance we chose the random forest model for evaluation on our held-out testing dataset. This model achieved an AUC of 65% to predict any adenoma among all diminutive rectosigmoid polyps.

Cary Cotton and Jared Sninsky

BIOS 635

Introduction

Current screening and surveillance colonoscopy guidelines recommend the resection and histopathologic analysis of all polypoid lesions. However, diminutive polyps carry an exceedingly low risk of dysplasia and their removal is of dubious clinical benefit. In fact, only 50% of diminutive polyps are neoplastic in histology[1], even less in the rectosigmoid, and the costs of diminutive polyp removal are estimated to be one billion dollars annually[2].

Over the last decade, a “diagnose and leave” approach has been proposed, in which diminutive polyps within the rectosigmoid that appear hyperplastic can be left. However, current ESGE/ASGE guidelines only endorse this practice using narrow band imaging by trained and certified endoscopists with rigorous photo-documentation and validated scales [3]. Barriers to this approach remain high, with 84% of endoscopists not using this method and 60% believing that an optical diagnosis approach is not feasible[4]. There is also concern that not correctly identifying the histology of diminutive polyps will affect follow up surveillance intervals.

The ASGE has proposed that any “diagnose and leave” modality for rectosigmoid diminutive polyps should provide a 90% or greater negative predictive value for adenomatous histology. A post hoc analysis, which considered all rectosigmoid diminutive polyps as hyperplastic without optical diagnosis, coincided with pathology-based surveillance recommendations in 89% of patients[2]. Recent studies have emerged showing promise for AI-driven image analysis of polyps to determine histology; but implementation is hindered by current endoscopic technology and a lack of available, clinically validated software[5].

We propose that a machine learning algorithm derived from a large real-world clinical endoscopy database, the Clinical Outcomes Research Initiative (CORI), with associated

pathology will accurately predict diminutive rectosigmoid polyp histology using a combination of clinical variables and concurrent endoscopic findings. For example, patient demographics as well as the size and location of concurrent polyps in the colon are likely highly predictive of rectosigmoid polyp pathology.

Methods

Patients

Adult patients, ages 18-85, undergoing screening or surveillance colonoscopy were identified within version four of the CORI database which spans 2008-2014. The CORI database was developed by the American Society for Gastrointestinal Endoscopy and consists of endoscopic exams performed nationally in outpatient care centers, private practices, universities, Veteran's hospitals, and general hospitals.

Inclusion in the analysis

Patients with any rectosigmoid diminutive (1-5mm) polyp pathology were included in the cohort undergoing routine average risk screening or surveillance colonoscopy. Only patients with complete colonoscopic exams were included. Patients who had multiple polyps collected within a single pathology jar were excluded.

Variables

The outcome was the presence of adenomatous pathology of a diminutive (≥ 5 mm) rectosigmoid polyp. Adenocarcinoma and sessile serrated polyps were included within the "adenoma" group; while uncommon they are also entities for which histologic identification could confer a clinical benefit. Any patient with at least one diminutive rectosigmoid adenoma was grouped into the "adenoma" outcome group. Non-adenomatous diminutive rectosigmoid polyps

included any pathology not considered premalignant, including hyperplastic polyps, normal colonic mucosa, and adipose tissue.

Predictors

The three main categories of predictor variables include: demographic features, colonoscopy indication details, and concurrent gross colonoscopy findings. Specifically, demographic variables included age, race, sex, BMI, and smoking/alcohol status. Colonoscopy indication variables included years from last colonoscopy and most advanced lesion on last colonoscopy. Concurrent colonoscopy findings were identified such as hemorrhoid characteristics, diverticulosis presence and location, as well as polyp size, shape (sessile vs pedunculated) and location.

Machine learning analysis

A machine learning classification analysis was performed with any adenomatous diminutive rectosigmoid polyp was analyzed as the binary outcome using the *caret* package within R [14]. The data was split *a priori* into training and testing sets using a 60/40 hold-out approach. Cross validation within the training hold-out set was used to tune the candidate machine learning models. The models investigated included logistic regression, naive bayes, random forest, and K nearest-neighbor. Using the Caret *TrainControl* function, the models were tuned individually for sensitivity due to the desire to optimize negative predictive value. Each model was also tuned with and without the use of SMOTE: synthetic minority oversampling technique. After tuning the models, the accuracy parameters of AUC, sensitivity, specificity, balanced accuracy, positive predictive value, and negative predictive value were calculated for each model within the cross validated training set.

Next, we identified the cut-off probability threshold based on a cost-benefit method in which the cost of one false negative equals five false positives. This was determined as the clinical cost of leaving in a hyperplastic polyp is negligible while there is a small but real risk malignant potential when leaving an adenomatous polyp. Negative predictive value at the specified cut-off was derived using the database study prevalence and extrapolated with the true known adenoma prevalence of all diminutive rectosigmoid polyps. The candidate model with the highest AUC, NPV, and percent sample ruled-out was chosen to test on the hold out testing dataset.

Results

There was an equal distribution of variables between patients who had at least one diminutive rectal adenoma and those with diminutive non-adenoma pathology (**table 1**). There was a 38% prevalence of patients with at least one rectosigmoid adenoma. No notable differences were observed comparing our training and testing data-sets (**table 2)**.

Tuning the models with repeated cross-validated in the training dataset yielded an AUC of 63%, 58%, 57%, and 65% for logistic regression with LASSO, naive bayes, KNN, and random forest respectively. None of the models had an increase in AUC after upsampling the minority class with SMOTE (**table 3**). Using the cost-benefit threshold parameter, we ran accuracy metrics on each model (**table 4**). Within the training dataset, the random forest model had a sensitivity of 41% , specificity of 79%, PPV 55%, and NPV 74% (**figure 2**). The random forest model applied to the hold-out testing set had an AUC of 0.65(**figure 3**) and a sensitivity of 71%, specificity of 51%, PPV 46%, and NPV 75% (**table 5**). XXX percent of the testing

patients were ruled out with this NPV. After adjusting for the known prevalence of diminutive rectosigmoid adenomas at 20%, the re-calculated NPV on the testing data was 88%.

Discussion and Conclusion

Studies have estimated that reducing unnecessary pathologic assessment of low-risk diminutive rectosigmoid polyps would exceed one billion dollars annually [6]. In order to adopt a “diagnose and leave” strategy, the ASGE recommends a negative predictive value of at least 90% for adenomatous polyps in the rectosigmoid. We used a large real-world endoscopic database, to predict the presence of any adenomatous rectosigmoid polyp, using demographic data and gross concurrent colonoscopic findings. Using a random forest model, on our held-out testing cohort, we achieved an ROC of 65 with a negative predictive value of 68% at a predetermined threshold. We propose that incorporating clinical data with image analysis with machine learning models, also known as multimodal machine learning, will lead to refined diagnostic accuracy within the field of gastrointestinal endoscopy.

Expert analysis of diminutive colorectal polyps with narrow band imaging has revealed a negative predictive value up to 92% [7]. However, studies involving non-expert endoscopists with NBI have demonstrated less satisfactory accuracy, often not fulfilling the PIVI criteria [8]. After a standardized NBI training program, Patel et al found a auROC of 75%, sensitivity 88%, and specificity 78% for diminutive rectosigmoid polyps; however, only 20/26 endoscopists met the 90% NPV threshold after training[9]. Recently, a deep convolutional neural network have been fed colonoscopic videos with NBI and trained to decipher between hyperplastic and adenomatous polyps[10].

In clinical practice, some diminutive rectosigmoid polyps are not resected that appear overtly hyperplastic to the endoscopist. Consequently, we believe that in our real-world endoscopic database there was a bias enriching the database with indeterminate lesions and adenomas. This is evidenced by the fact that we observed a 39% prevalence of diminutive rectosigmoid adenomas while prospective studies have noted a prevalence of 20-25%[11] Interestingly, after adjusting for the real-word prevalence of 20%, our NPV for our given threshold was 88%, effectively ruling out XXX% of patients.

Our study demonstrated that, with only the inclusion of limited clinical data and concurrent gross endoscopic findings, we predict the presence of any diminutive rectosigmoid polyp with a reasonable degree of accuracy. Just as clinicians synthesize image analysis within a clinical context, artificial intelligence modalities are increasingly merging clinical metadata with imaging and video analysis [12].This practice is known as multimodal deep learning and was first used in automated vehicles merging LiDaR and light detection systems , but very recently has been utilized in diagnostic models for Alzheimer's disease[13]. The incorporation of key clinical variables may have the potential to boost video-based optical diagnosis artificial intelligence.

Based on rudimentary demographic and endoscopic findings, without any imaging or video data, we were able to achieve a negative predictive value of 74% for diminutive rectosigmoid adenomas in a held-out testing cohort, 88% after adjusting for prevalence. Artificial intelligence is increasingly applied to gastroenterology, ranging from video polyp detection to cross-sectional imaging analysis. We have demonstrated that clinical/endoscopic data alone can lead to surprisingly accurate diagnostic models and merging these models with other modalities will further improve diagnostic accuracy and clinical care in gastroenterology.

References

1. von Renteln, D. and H. Pohl, *Polyp Resection - Controversial Practices and Unanswered Questions.* Clinical and translational gastroenterology, 2017. **8**(3): p. E76-e76.
2. Rex, D.K., *Risks and potential cost savings of not sending diminutive polyps for histologic examination.*Gastroenterology & hepatology, 2012. **8**(2): p. 128-130.
3. Abu Dayyeh, B.K., et al., *ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps.*Gastrointestinal Endoscopy, 2015. **81**(3): p. 502.e1-502.e16.
4. Willems, P., et al., *Uptake and barriers for implementation of the resect and discard strategy: an international survey.* Endosc Int Open, 2020. **8**(5): p. E684-e692.
5. Zachariah, R., et al., *Prediction of Polyp Pathology Using Convolutional Neural Networks Achieves "Resect and Discard" Thresholds.* Am J Gastroenterol, 2020. **115**(1): p. 138-144.
6. Kessler, W.R., et al., *A quantitative assessment of the risks and cost savings of forgoing histologic examination of diminutive polyps.* Endoscopy, 2011. **43**(8): p. 683-91.
7. Kaltenbach, T., et al., *Real-time optical diagnosis for diminutive colorectal polyps using narrow-band imaging: the VALID randomised clinical trial.* Gut, 2015. **64**(10): p. 1569.
8. Ladabaum, U., et al., *Real-time optical biopsy of colon polyps with narrow band imaging in community practice does not yet meet key thresholds for clinical decisions.* Gastroenterology, 2013. **144**(1): p. 81-91.
9. Patel, S.G., et al., *Real-Time Characterization of Diminutive Colorectal Polyp Histology Using Narrow-Band Imaging: Implications for the Resect and Discard Strategy.* Gastroenterology, 2016. **150**(2): p. 406-18.
10. Byrne, M.F., et al., *Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model.* Gut, 2019. **68**(1): p. 94.
11. Wallace, M.B., et al., *Accuracy of in vivo colorectal polyp discrimination by using dual-focus high-definition narrow-band imaging colonoscopy.* Gastrointest Endosc, 2014. **80**(6): p. 1072-87.

12. Huang, S.C., et al., *Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines.* NPJ Digit Med, 2020. **3**: p. 136.
13. Thung, K.H., P.T. Yap, and D. Shen, *Multi-stage Diagnosis of Alzheimer's Disease with Incomplete Multimodal Data via Multi-task Deep Learning.* Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2017), 2017. **10553**: p. 160-168.
14. R Core Team (2020). R: *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Cary Cotton and Jared Sninsky

Tables

Table 1. Demographic, clinical, and endoscopic characteristics of patients with an adenoma among rectosigmoid polyps compared to patients without an adenoma.

	Diminutive Rectosigmoid Pathology			
Characteristic	**Adenoma, N = 2,670[1]**	**Nonadenoma, N = 4,330[1]**	**p-value[2]**	**Missing (N)**
Age	60.8 (8.7)	59.5 (8.3)	<0.001	15
Race/ethnicity			<0.001	114
White	2,310 / 2,627 (87.9%)	3,594 / 4,259 (84.4%)		
Black	147 / 2,627 (5.6%)	410 / 4,259 (9.6%)		
Asian	150 / 2,627 (5.7%)	205 / 4,259 (4.8%)		
Hispanic	20 / 2,627 (0.8%)	50 / 4,259 (1.2%)		
BMI class			0.2	6,504
Normal	30 / 150 (20.0%)	56 / 346 (16.2%)		
Obese	59 / 150 (39.3%)	168 / 346 (48.6%)		
Overweight	60 / 150 (40.0%)	119 / 346 (34.4%)		
Underweight	1 / 150 (0.7%)	3 / 346 (0.9%)		
Alcohol use			0.007	6,268
Abstains	96 / 247 (38.9%)	219 / 485 (45.2%)		
Heavy (>2 drinks per day)	15 / 247 (6.1%)	10 / 485 (2.1%)		
Moderate (1-2 drinks per day)	30 / 247 (12.1%)	39 / 485 (8.0%)		
Occasional (average less than daily)	106 / 247 (42.9%)	217 / 485 (44.7%)		
Polyp number in ascending colon	0.3 (0.6)	0.2 (0.5)	<0.001	0
Polyp number in cecum	0.1 (0.4)	0.1 (0.3)	<0.001	0
Polyp number in descending colon	0.2 (0.5)	0.1 (0.4)	<0.001	0
Polyp number in rectum	0.5 (0.7)	0.7 (0.8)	<0.001	0
Polyp number in sigmoid colon	0.9 (0.9)	0.7 (0.8)	<0.001	0
Diverticulosis	1,371 / 2,667 (51.4%)	2,009 / 4,320 (46.5%)	<0.001	13
Hemorrhoids			<0.001	995
None	1,295 / 2,278 (56.8%)	2,309 / 3,727 (62.0%)		
Mild	520 / 2,278 (22.8%)	797 / 3,727 (21.4%)		
Moderate	416 / 2,278 (18.3%)	559 / 3,727 (15.0%)		
Severe	47 / 2,278 (2.1%)	62 / 3,727 (1.7%)		

[1]Mean (SD); n / N (%)

[2]One-way ANOVA; Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates); Fisher's exact test

Table 2. Comparing characteristics between the training and held out test datasets.

	Diminutive Rectosigmoid Pathology			
Characteristic	**test, N = 2,799[1]**	**train, N = 4,201[1]**	**p-value[2]**	**Missing (N)**
Age	59.9 (8.5)	60.1 (8.5)	0.4	15
Race/ethnicity			0.13	114
White	2,341 / 2,753 (85.0%)	3,563 / 4,133 (86.2%)		
Black	219 / 2,753 (8.0%)	338 / 4,133 (8.2%)		
Asian	161 / 2,753 (5.8%)	194 / 4,133 (4.7%)		
Hispanic	32 / 2,753 (1.2%)	38 / 4,133 (0.9%)		
BMI class			0.4	6,504
Normal	37 / 196 (18.9%)	49 / 300 (16.3%)		
Obese	81 / 196 (41.3%)	146 / 300 (48.7%)		
Overweight	76 / 196 (38.8%)	103 / 300 (34.3%)		
Underweight	2 / 196 (1.0%)	2 / 300 (0.7%)		
Alcohol use			0.3	6,268
Abstains	140 / 297 (47.1%)	175 / 435 (40.2%)		
Heavy (>2 drinks per day)	10 / 297 (3.4%)	15 / 435 (3.4%)		
Moderate (1-2 drinks per day)	27 / 297 (9.1%)	42 / 435 (9.7%)		
Occasional (average less than daily)	120 / 297 (40.4%)	203 / 435 (46.7%)		
Polyp number in ascending colon	0.2 (0.5)	0.2 (0.6)	0.2	0
Polyp number in cecum	0.1 (0.4)	0.1 (0.3)	0.3	0
Polyp number in descending colon	0.2 (0.5)	0.2 (0.5)	0.7	0
Polyp number in rectum	0.6 (0.8)	0.6 (0.8)	0.9	0
Polyp number in sigmoid colon	0.8 (0.8)	0.8 (0.8)	0.3	0
Diverticulosis	1,347 / 2,791 (48.3%)	2,033 / 4,196 (48.5%)	0.9	13
Hemorrhoids			0.11	995
None	1,443 / 2,427 (59.5%)	2,161 / 3,578 (60.4%)		
Mild	556 / 2,427 (22.9%)	761 / 3,578 (21.3%)		
Moderate	394 / 2,427 (16.2%)	581 / 3,578 (16.2%)		
Severe	34 / 2,427 (1.4%)	75 / 3,578 (2.1%)		
Histology			0.4	0
Adenoma	1,052 / 2,799 (37.6%)	1,618 / 4,201 (38.5%)		
Nonadenoma	1,747 / 2,799 (62.4%)	2,583 / 4,201 (61.5%)		

[1]Mean (SD); n / N (%)

[2]One-way ANOVA; Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates); Fisher's exact test

Table 3. Performance measures by five-repeated, five-fold cross-validation for the four models fit in training data with and without resampling using the SMOTE procedure.

Prediction model	Resampled with SMOTE	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	AUC
Lasso Regularized Logistic Regression	No	0.290 (0.008)	0.85 (0.003)	0.55 (0.007)	0.66 (0.002)	0.63 (0.003)
	Yes	0.420 (0.003)	0.75 (0.003)	0.51 (0.002)	0.67 (0.000)	0.62 (0.002)
K nearest neighbors	No	0.320 (0.005)	0.77 (0.003)	0.46 (0.003)	0.64 (0.001)	0.57 (0.003)
	Yes	0.400 (0.006)	0.68 (0.011)	0.44 (0.007)	0.64 (0.003)	0.56 (0.007)
Naive Bayes	No	0.340 (0.015)	0.78 (0.017)	0.49 (0.008)	0.65 (0.001)	0.58 (0.003)
	Yes	0.720 (0.017)	0.33 (0.032)	0.41 (0.013)	0.66 (0.027)	0.56 (0.018)
Random Forest	No	0.410 (0.002)	0.79 (0.005)	0.55 (0.007)	0.68 (0.002)	0.65 (0.004)
	Yes	0.400 (0.007)	0.77 (0.007)	0.53 (0.008)	0.67 (0.003)	0.63 (0.005)

Table 4. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) at the selected threshold in the training data as well as calculated NPV assuming a 20% population prevalence of adenoma.

Prediction model	SMOTE	Sensitivity at cutpoint	Specificity at cutpoint	PPV at cutpoint	NPV at cutpoint	NPV at 20% adenoma	Proportion ruled out
Lasso Regularized Logistic Regression	No	0.73 (0.03)	0.46 (0.03)	0.46 (0.01)	0.73 (0.01)	0.87 (0.00)	0.39 (0.03)
	Yes	0.69 (0.02)	0.48 (0.02)	0.45 (0.00)	0.71 (0.01)	0.86 (0.00)	0.41 (0.02)
K nearest neighbors	No	0.63 (0.00)	0.47 (0.01)	0.42 (0.00)	0.67 (0.00)	0.83 (0.00)	0.43 (0.01)
	Yes	0.78 (0.04)	0.30 (0.04)	0.41 (0.00)	0.68 (0.01)	0.84 (0.01)	0.27 (0.04)
Naive Bayes	No	0.57 (0.04)	0.56 (0.05)	0.45 (0.01)	0.68 (0.00)	0.84 (0.00)	0.51 (0.05)
	Yes	0.51 (0.14)	0.60 (0.14)	0.44 (0.02)	0.66 (0.02)	0.83 (0.01)	0.56 (0.14)
Random Forest	No	0.73 (0.05)	0.46 (0.07)	0.46 (0.01)	0.74 (0.01)	0.88 (0.01)	0.39 (0.06)
	Yes	0.72 (0.05)	0.46 (0.05)	0.46 (0.01)	0.73 (0.01)	0.87 (0.01)	0.39 (0.05)

Table 5. Two-by-two table, sensitivity, specificity, positive predictive value, and negative predictive value in the reserved test set.

	Reference	
Prediction	Adenoma	Nonadenoma
Adenoma	705	813
Nonadenoma	283	854

	Value
Sensitivity	0.71
Specificity	0.51
Pos Pred Value	0.46
Neg Pred Value	0.75

Figures

Figure 1. Receiver-operator characteristic curves for the fit models under repeated cross-validation in the training data across the five missing data imputations.

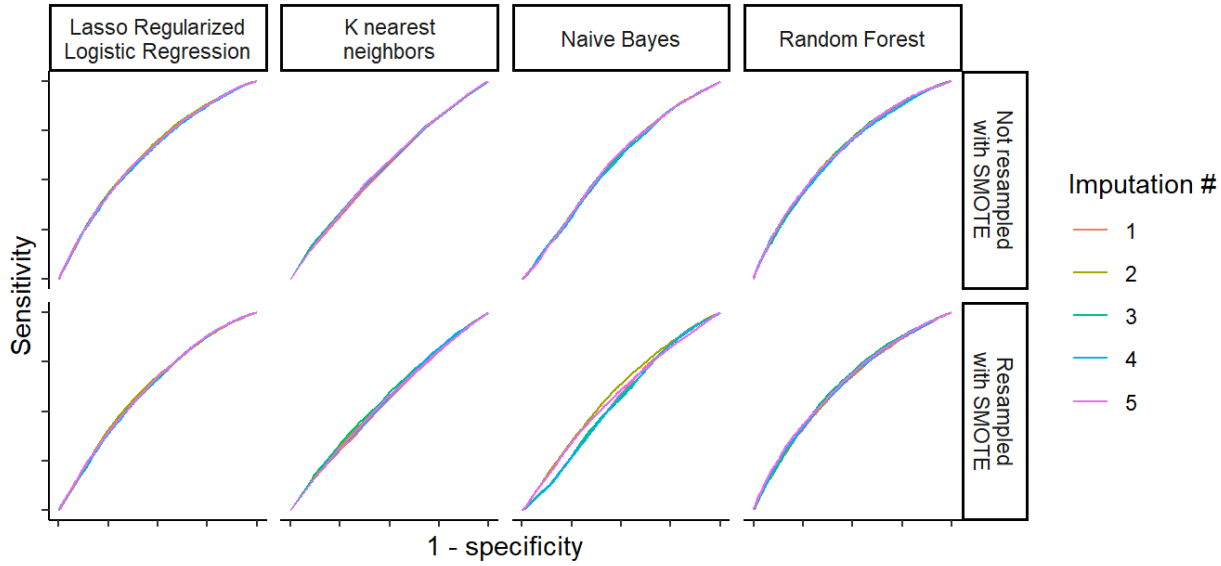

Figure 2. Threshold selection in the selected random forest model without SMOTE resampling using maximum kappa weighting false negatives five times higher than false positives.

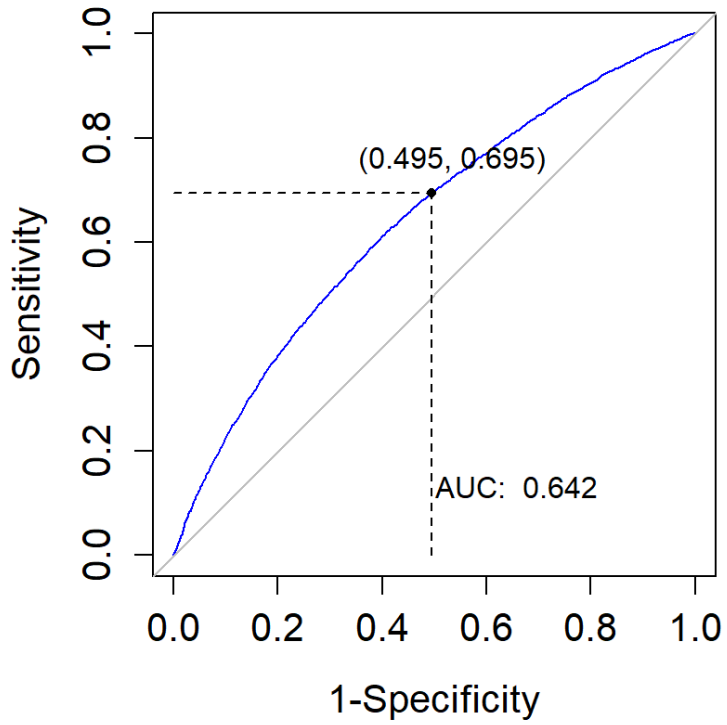

Figure 3. Receiver-operator characteristic (ROC) curve in the testing data and area under the ROC curve (AUC).