IFSO

ORIGINAL CONTRIBUTIONS

Development and Validation of an Artificial Intelligence-Based Model to Predict Gastroesophageal Reflux Disease After Sleeve Gastrectomy

Sameh Hany Emile[1] · Waleed Ghareeb[2] · Hossam Elfeki[1] · Mohamed El Sorogy[3] · Amgad Fouad[3] · Mohamed Elrefai[3]

Abstract

Purpose Prediction of the onset of de novo gastroesophageal reflux disease (GERD) after sleeve gastrectomy (SG) would be helpful in decision-making and selection of the optimal bariatric procedure for every patient. The present study aimed to develop an artificial intelligence (AI)-based model to predict the onset of GERD after SG to help clinicians and surgeons in decision-making.

Materials and Methods A prospectively maintained database of patients with severe obesity who underwent SG was used for the development of the AI model using all the available data points. The dataset was arbitrarily split into two parts: 70% for training and 30% for testing. Then ranking of the variables was performed in two steps. Different learning algorithms were used, and the best model that showed maximum performance was selected for the further steps of machine learning. A multitask AI platform was used to determine the cutoff points for the top numerical predictors of GERD.

Results In total, 441 patients (76.2% female) of a mean age of 43.7 ± 10 years were included. The ensemble model outperformed the other models. The model achieved an AUC of 0.93 (95%CI 0.88–0.99), sensitivity of 79.2% (95% CI 57.9–92.9%), and specificity of 86.1% (95%CI 70.5–95.3%). The top five ranked predictors were age, weight, preoperative GERD, size of orogastric tube, and distance of first stapler firing from the pylorus.

Conclusion An AI-based model for the prediction of GERD after SG was developed. The model had excellent accuracy, yet a moderate sensitivity and specificity. Further prospective multicenter trials are needed to externally validate the model developed.

Keywords Artificial intelligence · Sleeve gastrectomy · Gastroesophageal reflux disease · Predict · Model

---

Key points
- 441 patients (76.2% female) of a mean age of 43.7 ± 10 years were included.
- The ensemble model outperformed the other models.
- The model achieved an AUC of 0.93, sensitivity of 79.2%, and specificity of 86.1%.
- The top-ranked predictors were age, weight, preoperative GERD, bougie size, and distance from the pylorus.

Sameh Emile and Waleed Gahreeb contributed equally to the study and are considered co-first authors.

✉ Sameh Hany Emile
sameh200@hotmail.com

Waleed Ghareeb
waleed.m.ghareeb@med.suez.edu.eg

Hossam Elfeki
hossamelfeki@gmail.com

Mohamed El Sorogy
sorogy@gmail.com

Amgad Fouad
amgfouad@yahoo.com

Mohamed Elrefai
mohelrefai@mans.edu.eg

1 General Surgery Department, Mansoura University Hospitals, Mansoura University, Mansoura, Egypt

2 Gastrointestinal Surgery Unit, Department of Surgery, Faculty of Medicine, Suez Canal University Hospital, Ismailia, Egypt

3 Gastrointestinal Surgery Centre, Mansoura University, Mansoura, Egypt

Introduction

Sleeve gastrectomy (SG) is the most commonly performed bariatric procedure across the world [1]. The popularity of SG has emanated from the excellent results attained by the procedure in terms of weight loss and improvement in comorbidities, in addition to its good safety profile [2]. However, akin to every other surgical procedure, SG is associated with a

number of adverse events, including staple line leak and bleeding, gastric stenosis, volvulus, and gastroesophageal reflux disease (GERD).

A recent meta-analysis of 46 studies comprising more than 10,000 patients found that the long-term prevalence of GERD and Barrett's esophagus after SG is 28% and 8%, respectively [3]. Moreover, 4% of patients may need conversion from SG to Roux-en-Y gastric bypass (RYGB) to treat severe reflux esophagitis. Several hypotheses have been suggested to explain post-SG GERD, and they include increased intraluminal pressure, loss of gastric compliance, removal of the gastric fundus, decreased lower esophageal sphincter (LES) pressure, twisting of the pouch, and persistence of hiatal hernia [4]. In contradiction, other reports documented an improvement in preoperative GERD after SG [5, 6]. Mechanisms involved in the relief of GERD after SG may include reduction of acid production, decrease in intra-abdominal pressure, reduction of gastric volume, and acceleration of gastric emptying [4].

These conflicting findings stimulated further investigation on the impact of SG on the anatomic and physiologic function of the gastroesophageal junction. A recent prospective study [7] concluded a significant decrease in the resting LES pressure after SG in patients without preoperative GERD, yet not in patients with symptomatic GERD. Also, esophageal manometry seemed to have a limited role in the detection of GERD after SG, unlike 24-h pH monitoring which could clearly identify improvement or persistence of GERD.

The prediction of the onset of de novo GERD after SG has become an utmost need to help guide decision-making and selection of the optimal bariatric procedure for every patient [8]. A study including 213 patients found that higher preoperative BMI was less likely to be associated with new-onset or worsening of GERD symptoms whereas more severe heartburn symptoms on standing were predictive of increased risks of developing or worsening GERD symptoms [9].

It has been assumed that if GERD can be predicted before SG, it would guide clinicians to either select another less refluxogenic procedure such as RYGB or to counsel the patients about the need to do frequent upper GI endoscopy after SG, should they still want to have the procedure. Therefore, the present study aimed to develop an artificial intelligence (AI)-based model to predict the onset of GERD after SG to help clinicians and surgeons in decision making.

Patients and Methods

Study Design and Setting

This was a retrospective review of prospectively maintained data on patients with severe obesity who underwent SG in the period of January 2011 through December 2019. Patients had their surgical treatment in the General Surgery Department and Gastrointestinal Surgery Center of Mansoura University. Ethical approval for the study was obtained from the institutional review board (IRB) of Mansoura Faculty of Medicine and the study was conducted in compliance with the ethical principles of the Declaration of Helsinki.

Patient Population

The study enrolled adult patients of either sex with severe obesity defined as BMI > 40 kg/m^2 or > 35 kg/m2 with at least one associated major comorbidity who underwent laparoscopic SG (LSG). Patients with secondary obesity due to endocrine disorders and those who underwent revisional bariatric procedures or bariatric operations other than SG were excluded. Patients with missing data on the outcome of SG and preoperative and postoperative GERD were also excluded.

Technique of Sleeve Gastrectomy

All procedures were done under general anesthesia with the patients placed in the French position. Procedures were performed or supervised by expert laparoscopic and bariatric surgeons. The procedure was done in a standardized stepwise manner as follows.

Pneumoperitoneum at 12 mm-Hg was created; then a standard four-port technique was used. Harmonic ace scalpel ™ was used to dissect the greater omentum off the greater.

curvature, beginning 2–6 cm away from the pylorus and proceeding up to the angle of His. Then, an orogastric tube (size of 33, 36, 38, or 42 Fr) was placed in the stomach for calibration followed by transection of the stomach using linear staplers, starting 2–6 cm from the pylorus up to the.

angle of Hiss. Upon creation of the gastric pouch, hemostasis was ascertained; then an intraperitoneal drain was placed before deflation of the abdomen. Hiatal defects were repaired routinely with suture cruroplasty during SG.

Assessments

Patients were assessed preoperatively with routine laboratory investigations, EKG, chest X-ray, upper GI endoscopy, and pulmonary function tests. Postoperatively, patients were assessed at one and two weeks then at 3, 6, and 12 months postoperatively. Follow-up upper GI endoscopy was performed as part of routine follow-up at 6 and 12 months after SG. Patients with symptoms of GERD before were assessed using the GERD-Q score [10]. A total score of 0 to 2 points implied 0% likelihood of GERD; 3 to 7 points implied 50% likelihood; 8 to 10 points implied 79% likelihood; 11 to 18 points implied 89% likelihood. Patients with a score greater

than 2 ($\geq 50\%$ likelihood of GERD) or otherwise with a known history of GERD were investigated with upper GI endoscopy.

Endoscopy identified GERD as single or multiple (circumferential or non-circumferential) erosions above the gastroesophageal junction and graded it according to the Savary-Miller classification into four grades. Patients with confirmed GERD were under proton pump inhibitor treatment. Weight loss was expressed as %total weight loss (TWL) and was calculated as [(preoperative weight – weight on follow-up)/preoperative weight]$\times 100$.

Data Collected

The following data were extracted from patients' records:

- Demographic data: age, sex, weight, BMI, and smoking status
- Medical comorbidities: diabetes mellitus (DM), hypertension, hyperlipidemia, sleep apnea, and preoperative GERD
- Technical details: size of orogastric calibration tube, distance of transection from the pylorus, staple line reinforcement, and intraoperative complications
- Weight loss at 6 and 12 months after SG
- Postoperative GERD as confirmed by upper GI endoscopy

Development of the AI-Based Model

Using the available data points, an AI-based model will be developed employing machine learning algorithms through the following steps.

Data Preprocessing

Patients' data were refined to make machine learning possible. The dataset was arbitrarily split into two parts: 70% for training and 30% for testing. The training dataset was validated five times to avoid the overfitting of the model and was randomly divided into five parts throughout the validation process, four parts for learning and one for validation.

Variable Selection for Model Creation

All variables were ranked and queued based on their importance to make the final prediction. Testing the estimate levels of importance aimed to minimize redundant variables while keeping the maximum relevance to the proposed prediction. Due to data type heterogeneity (numerical and categorical variables), ranking of the variables was performed in two steps: the first step was done using the interaction test algorithm while the second step was permuting out-of-bag

observations (variables). The permuted variables estimates were not biased toward numerical predictors (which contain many levels). Estimates of greater importance indicated more important predictors. For accurate selection of important predictors, estimating the predictor importance by permuting out-of-bag observations was compared to the estimates obtained by summing gains in the mean squared error (MSE).

Machine Learning Process

Different learning algorithms were used, and the best model that showed maximum performance was selected for the further steps of machine learning. The performance was judged based on the accuracy rate and the area under the receiver operator curve (AUC, ROC). Furthermore, the developed model underwent an optimization process to minimize the predicting error. The desktop application for the GERD prediction AI model can be downloaded from the following link: https://www.dropbox.com/sh/z1nqd7jtpmp97z2/AADhgez0bYtkh-OqixSl0MMca?dl=0.

Statistical Analysis

Analysis of data was conducted using SPSS program (version 23, IBM) for Windows. Continuous variables were described as mean and standard deviation (SD) or median and normal range if data were not normally distributed. Student's *t* test was used to compare continuous variables. Categorical variables were expressed as numbers and proportions and were compared using Chi-square or Fisher exact test. A *p* value less than 0.05 was considered significant.

Results

Characteristics of Patients

The present study included 441 patients of a mean age of 43.7 ± 10 years. Patients were 336 (76.2%) female and 105 (23.8%) male. The mean weight of patients was 136.3 ± 25.1 kg, and the mean BMI was 50.7 ± 7.7 kg/m^2. In total, 44 (9.9%) patients had DM, 77 (17.5%) had hypertension, 95 (21.5%) had sleep apnea, and 35 (7.9%) had hyperlipidemia. Nine patients were smokers (Table 1).

A total of 91 (20.6%) patients had preoperative GERD with a median GERD-Q score of 9 (range, 3–16). By endoscopic assessment, 84.6% of patients had grade I GERD, 11% had grade II GERD, and 4.4% had grade III GERD. A total of 18 patients had evidence of hiatus hernia in endoscopy and intraoperatively.

Technical Details

A 33-F orogastric tube was used in 4 (0.9%) patients, 36 F in 184 (41.7%) patients, 38 F in 230 (52.1%) patients, and 42 F in one (0.2%) patient. Staple line reinforcement was done in 14 (3.2%) patients. The distance of starting gastric transection from the pylorus varied from 2 cm (123 patients), 3 cm (14 patients), 4 cm (194 patients), 5 cm (13 patients), and 6 cm (58 patients).

Weight Loss and Postoperative GERD

BMI decreased significantly from a baseline of 50.7 ± 7.7 to 34.1 ± 4.6 kg/m^2 at 12 months after SG. The mean %TWL at 12 months was 30.4 ± 6.9%.

Among 91 patients with preoperative GERD, 50 (55%) showed improvement or resolution in GERD after SG whereas 41 (45%) had persistent or worsening GERD symptoms. In addition, 46 additional patients developed new-onset GERD after SG, amounting to a total of 87 (19.7%) patients who had postoperative GERD (49.4% grade I, 38% grade II, and 11.5% grade III).

Selection of GERD Predictors

The study cohort has been divided into 90% for machine learning set (training set = 70% and validation set = 20%) and 10% for testing set. There was no statistical difference between the input predicting variables of the two sets (Table 2). Predictor selection minimizes the amount of data used in machine learning by choosing only the most important predictors/variables that give the best accuracy for the prediction process. The included predictors were ranked by importance to predict GERD after SG. As an initial step to check for predictors' importance, all predictors were deemed important, and none was excluded. Preoperative GERD, distance from the pylorus, orogastric tube size, intraoperative complication, and age were ranked as the top five important

Table 1 Characteristics of the cohort studies

Parameter	Value
Age	43.7 ± 10
Male/female	105/336
Mean weight in kg	136.3 ± 25.1
Mean body mass index in kg/m^2	50.7 ± 7.7
Comorbidities	
Diabetes mellitus	44 (9.9%)
Hypertension	77 (17.5%)
Hyperlipidemia	35 (7.9%)
Sleep apnea	95 (21.5%)
GERD	91 (20.6%)
Smokers	9

predictors (Fig. 1A). Since there were two types of predictor types (continuous and categorical), a further algorithmic assessment was required to avoid bias of predictor selection by comparing the importance estimates to the mean squared error (MSE). Although all predictors remained important for the accurate prediction of GERD, the rank of predictors was permuted by importance (e.g., preoperative GERD, bougie size, age, weight, and distance from the pylorus) (Fig. 1B).

Machine Learning and Model Optimization

Five models were trained using 15 different algorithms. The ensemble model trained by the bagged algorithm outperformed the 15 models trained by different algorithms (Table 3). To ensure the quality of model performance, the ensemble model was tuned by a hyperparameter optimization process which was continued until reaching the slightest minimum classification error (0.097).

Visualization of Prediction/Decision-Making Process

Parallel coordinate plotting was used to visualize the predictors while contributing to the prediction/decision-making process. The predictors were arranged into descending order depending on the importance estimates shown in Fig. 1B. The parallel coordinates plot helped know how the model thinks while taking every input predictor until making its final decision. While the plot showed that the model failed to predict GERD when interpreting the first six important input predictors, inclusion of further predictors served to improve the accuracy of the model confusion.

Performance Assessment

The ensemble model achieved an AUC of 0.97 (95% CI 0.96–0.98) during the validation process. On assessing the performance on the testing dataset, the model achieved an AUC of 0.93 (95% CI 0.88–0.99) and AUCPR of 0.88 (95% CI 0.86% to 0.93%). The model had a sensitivity of 79.2% (95% CI 57.9% to 92.9%), specificity of 86.1% (95% CI 70.5% to 95.3%), PPV of 85.1% (95% CI 71.1% to 92.9%), and NPV of 80.5% (95% CI 65.2% to 90.1%) (Fig. 2).

Identifying Cutoff Values of GERD Predictors Using a Multitask AI Platform

After the prediction of risk to develop GERD after SG as the first task of the AI platform, patients who were classified as at high risk to develop GERD were included in the next step. The orogastric tube size and distance from the pylorus were two modifiable predictors and were significantly correlated to the risk of developing GERD ($p = 0.001$, 0.02. respectively). Out of 15 trained models with different

Table 2 Predictive variables of the machine learning and testing sets used in building predictive models for postoperative gastroesophageal reflux disease

Predictors	Machine learning set	Testing set	p value
Patients' age in years, mean (SD)	34.6 (9.8)	36.1 (11.7)	0.35
Patients' sex (%)			
Male	94 (23.7)	11 (25)	0.85
Female	303 (76.3)	33 (75)	
Weight (*Kg*), mean (SD)	135.9 (34.5)	140.1 (30.1)	0.31
BMI (*Kg/m²*), mean (SD)	50.7 (7.6)	51.3 (9.4)	0.64
Diabetes mellitus (%)			
Yes	41 (10.3)	3 (10)	0.46
No	356 (89.7)	41 (90)	
Hypertension (%)			
Yes	66 (16.6)	11 (17.5)	0.17
No	331 (83.4)	33 (82.5)	
Preop GERD (%)			
Yes	83 (20.9)	8 (18.2)	0.74
No	313 (79.1)	36 (81.8)	
Smoking (%)			
Yes	8 (2)	1 (2.3)	0.91
No	389 (98)	43 (97.7)	
Bougie size			
33 F	4(1.1)	0	0.78
36 F	168 (44.4)	16 (39)	
38 F	205 (54.2)	25 (61)	
42 F	1 (0.3)	0	
Staple line reinforcement (%)			
Yes	13 (3.3)	1 (2.3)	0.72
No	384 (96.7)	43 (97.7)	
Intraoperative complications (%)			
Yes	52 (13.1)	5 (11.3)	0.75
No	345 (86.9)	39 (88.7)	
Distance from the pylorus in cm, Mean (SD)	3.7 (1.3)	3.6 (1.5)	0.91
Postoperative GERD (%)			
Yes	79 (19.9)	8 (18.2)	0.79
No	318 (80.1)	36 (81.8)	

BMI, body mass index; GERD, gastroesophageal reflux disease
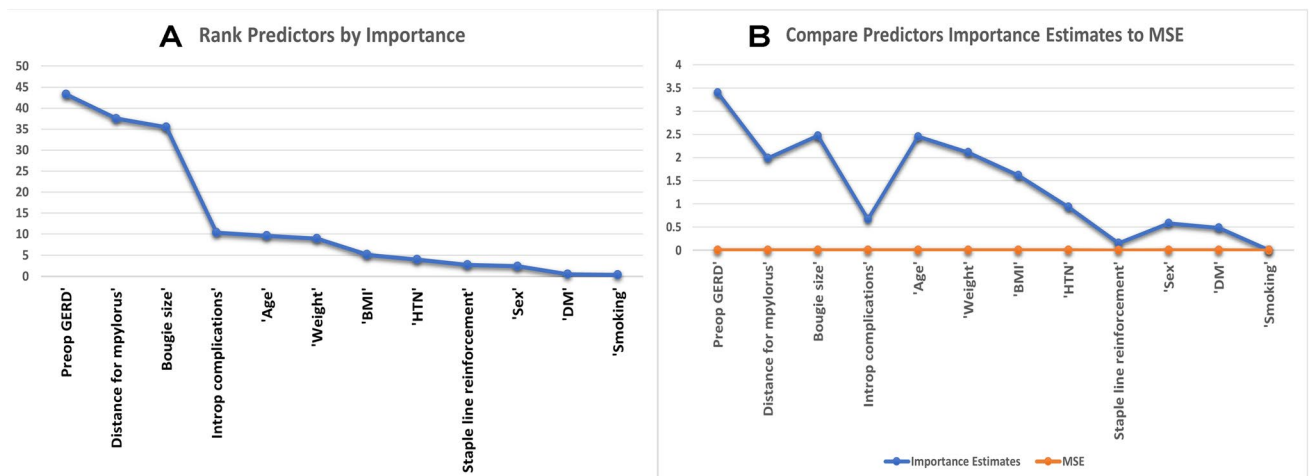

Fig. 1 **A** Ranking predictors by importance; **B** comparing predictor importance estimates

Table 3 Predictive performance of different models trained on different algorithms

Algorithm	AUC	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Decision tree					
Fine	0.84	78.48	78.93	78.83	78.57
Medium	0.8	60.09	84.28	79.26	67.86
Coarse	0.73	85.45	70.07	74.06	82.81
Logistic regression	0.76	68.57	69.62	69.30	68.90
Naïve Bayes					
Gaussian	0.75	65.92	70.75	69.27	67.49
Kemel	0.74	69.51	66.04	67.18	68.41
SVM					
Linear	0.73	43.50	87.42	77.57	60.74
Quadratic	0.8	57.40	84.28	78.50	66.42
Cubic	0.85	70.85	80.82	78.69	73.49
Fine	0.87	73.09	89.94	87.90	76.97
Medium	0.8	47.09	93.08	87.19	63.76
Coarse	0.76	45.74	89.94	81.97	62.37
Ensemble					
Boosted	0.85	73.09	83.33	81.43	75.59
Bagged	0.96	95.52	86.48	87.60	95.07
RUSBoosted	0.86	76.68	77.67	77.45	76.91

PPV, positive predictive value; NPV, negative predic0tive value, POD, postoperative day. [*]Statistically significant different AUC from the reference diagonal line at p value < 0.05

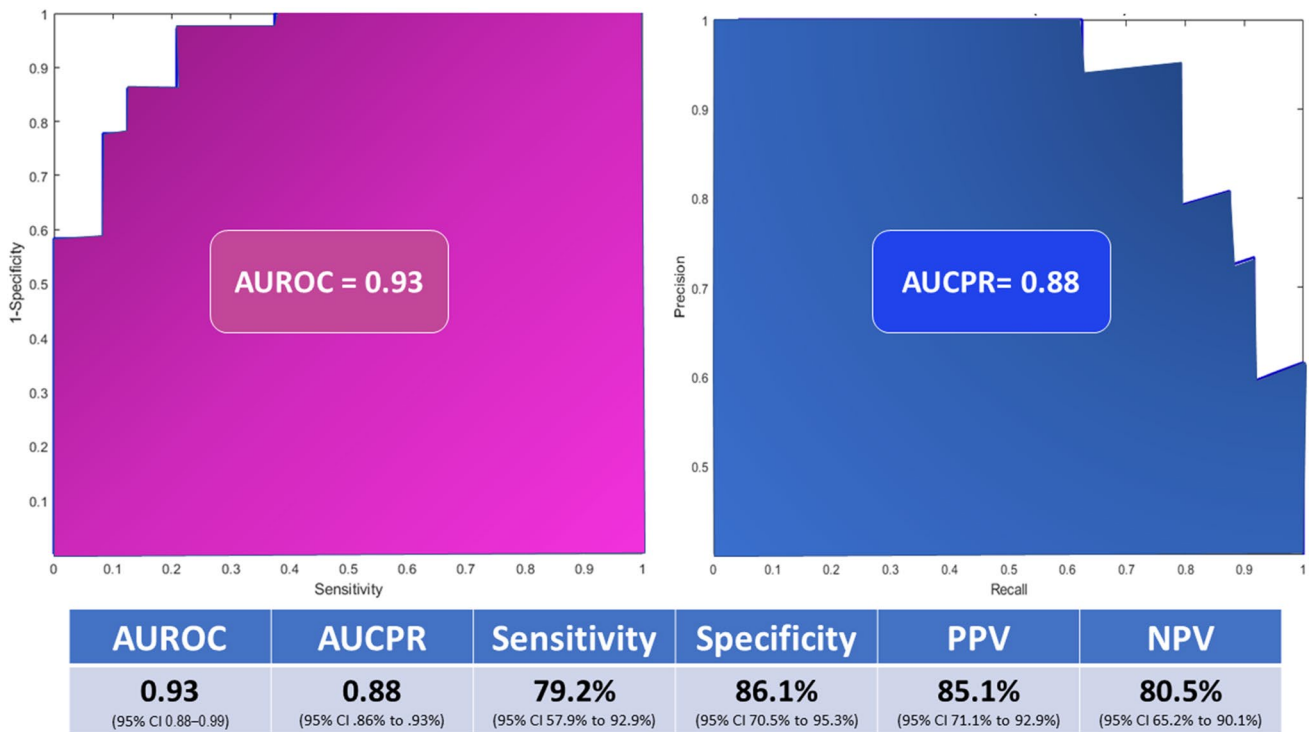| AUROC	AUCPR	Sensitivity	Specificity	PPV	NPV
0.93	**0.88**	**79.2%**	**86.1%**	**85.1%**	**80.5%**
(95% CI 0.88–0.99)	(95% CI .86% to .93%)	(95% CI 57.9% to 92.9%)	(95% CI 70.5% to 95.3%)	(95% CI 71.1% to 92.9%)	(95% CI 65.2% to 90.1%)

Fig. 2 Diagnostic accuracy parameters of the AI model developed

algorithms, the ensemble model trained by the RUSBoosted method showed the best performance (AU = 0.95) (Fig. 3A). The model predicted the orogastric size to be associated with a higher risk of GERD with an AUC of 0.9 (95% CI 0.86–0.94) and AUCPR of 0.76 (95% CI 0.71% to 0.83%). The third task of the AI platform was achieved by the support vector machine (SVM) model that showed the lowest root mean square error (RMSE 0.57) during machine

learning using the coarse Gaussian method (Fig. 3B). The model predicted the distance from the pylorus associated with a higher risk of GERD in the testing dataset with RMSE of 0.81.

The risk of developing postoperative GERD was remarkably increased at the age > 42 years, weight > 140.1 kg, BMI > 52.1 kg/m^2, orogastric tube < 38 Fr, and distance from the pylorus < 3 cm (odds ratio 1.03, 1.02, 1.1, 1.25, respectively) (Table 4).

A Sankey diagram illustrating how the surgeon's decision in choosing the bougie size and the distance from the pylorus can increase or decrease the risk of postoperative GERD in patients who have or do not have preoperative GERD can be accessed through this link file:///C:/Users/DELL/Downloads/Sankey.html.

A summary of the overall process of predictor selection and model development is shown in Fig. 4.

Fig. 3 Accuracy of different models used to determine which orogastric tube size and distance from the pylorus are associated with the greatest risk of gastroesophageal reflux disease after sleeve gastrectomy
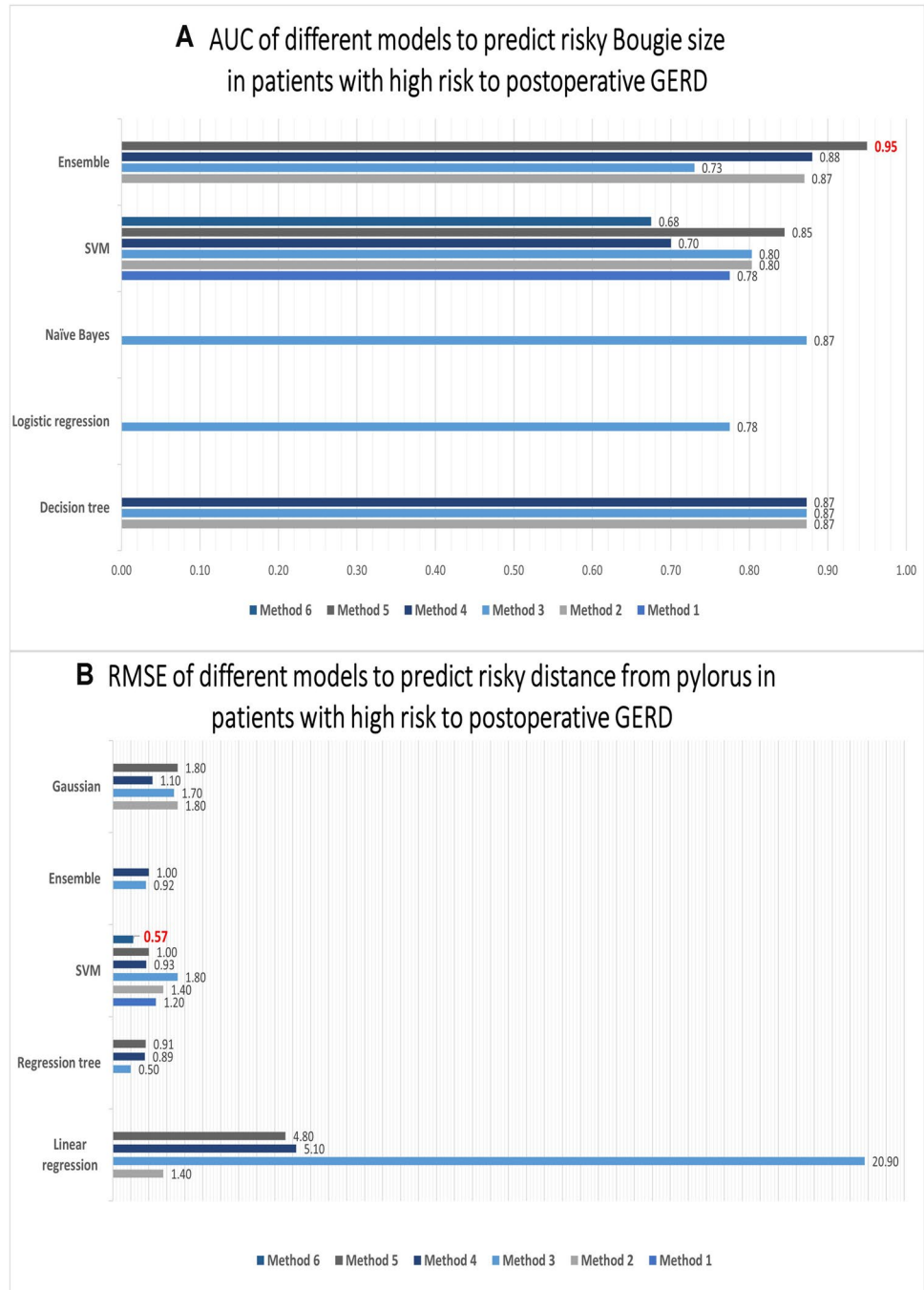
Table 4 Cutoff values and odds ratio of numerical predictors

Variables	Odds ratio	95% CI	Cutoff values
Age (*years*)	1.03	0.94–0.99	42
Weight (*Kg*)	1.02	0.99–1.04	140.1
BMI (*Kg/m²*)	1.1	0.86–0.98	52.1
Distance from the pylorus	1.25	0.66–0.98	3
(*cm*)	1.74	1.39 -2.18	38
Orogastric tube size (*Fr*)			

Discussion

SG has proven effective in the treatment of severe obesity and its associated comorbidities; however, this may come at the cost of increased risk of adverse events, particularly GERD. It has been estimated that SG may increase the odds of development of new-onset GERD by five times, as compared to other bariatric procedures such as RYGB [11]. This increased risk of GERD may also be associated with more grave consequences such as the development of Barrett's esophagus and potential onset of adenocarcinoma [3, 12].

Although there is a growing body of evidence that affirms the refluxogenic nature of SG, some reports documented an improvement in pre-existing GERD after SG [4]. Such a controversy implies different mechanisms by which SG impacts the gastroesophageal junction

physiology and gastric compliance. Therefore, it became necessary to predict which patients are more liable to develop new-onset GERD after SG which will help counsel these patients about this risk and serve to guide clinicians in choosing alternative procedures.

A number of predictive models for GERD have been described in the literature. A multivariate analysis model found that the visual detection of esophagitis on endoscopy and biopsy-proven esophagitis were strong independent predictors of GERD after SG. Moreover, female patients were more likely to develop GERD postoperatively as compared to male patients [13]. Another study assessed the predictive factors of GERD after bariatric surgery, comparing SG with RYGB. The study found that SG itself was the strongest independent predictor of de-novo GERD and that age, preoperative esophagitis, and postoperative hiatal hernia also independently predicted the onset of GERD after SG [14]. Further evidence was generated from a study [15] that included 213 patients who underwent SG, 47% of whom developed new-onset heartburn after SG. Patients with a higher baseline BMI had lower odds to develop new-onset or worsening symptoms of GERD while more severe heartburn symptoms on standing were associated with higher odds of GERD development ($OR = 1.22$).

It is worthy to note that all the previous predictive models were statistical models, based on multivariate logistic regression analyses. Although this type of models may be able to provide useful data to guide practice, it
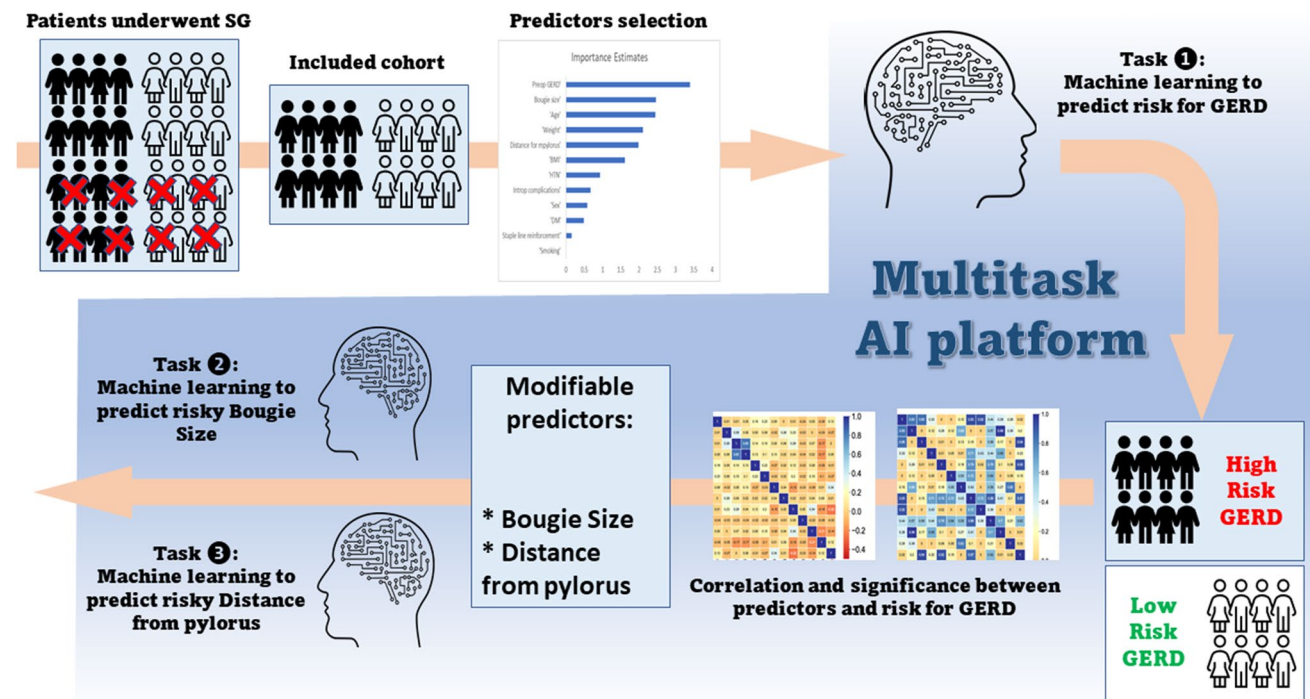

Fig. 4 Visual summary of the process of predictor selection and development of the model

is limited by the need for a large sample of data to avoid high standard errors in order to provide meaningful results [16]. Moreover, using the stepwise approach for multivariable regression may result in instability of the model and renders the model sensitive to slight changes in data such that addition or omission of a few observations can drastically alter the model [17].

The present model is based on a machine-learning AI model, the use of this kind of modeling has expanded widely in the last few years. AI has been used in a multitude of medical and surgical indications, including the differentiation of benign and malignant skin lesions, predicting colorectal anastomotic leak, and prediction of surgical complications in patients undergoing major abdominal surgery [18–20]. Recently, our group developed an AI-based model that was able to predict acute appendicitis with a sensitivity and accuracy that surpassed the traditional Alvarado scoring system [21].

Based on the initial promising results of AI-based predictive models, we decided to integrate baseline data of patients undergoing SG into an AI model, aiming to help predict the onset of GERD after surgery. Despite the average sensitivity of the model developed by our study (79.2%), the model had an excellent accuracy with an AUC of 97%. Hopefully, with further training and expanded learning process, the sensitivity of the AI model would increase.

Although the AI model included all the available baseline parameters, the top five ranked variables were the presence of baseline GERD, age, weight, size of orogastric tube, and distance of first stapler firing from the pylorus. These parameters, when present, would predictably impact the likelihood of postoperative GERD substantially. Age and BMI have been already recognized as independent predictors of post-SG GERD in previous studies [14, 15]. The presence of baseline GERD symptoms has been also identified as a strong predictor of postoperative GERD [22]. The AI model could recognize the three patient-related parameters among the top predictors of GERD. While older age may be expected to be associated with higher odds of postoperative GERD owing to the documented effects of aging on esophageal and esophagogastric junction mechanophysical properties [23], the association between baseline BMI and postoperative GERD seems to be contentious. Higher baseline BMI could be associated with a higher risk of GERD as it is independently associated with increased intragastric pressure which in consequence may cause an increase in the gastroesophageal-pressure gradient during inspiration, as de Vries and coworkers explained [24]. In contrast, another study [15] found higher baseline BMI to predict a lower incidence of GERD after SG. However, as the authors acknowledged, this finding should be interpreted with caution owing to the small magnitude of effects observed.

Two technical factors were among the top predictors of postoperative GERD as per the developed AI model. The size of orogastric tube reflects the size of the remaining gastric pouch, and with a smaller tube size, a smaller pouch would be constructed. It has been noted that smaller pouches are associated with a higher intraluminal pressure which eventually increases the risk of GERD. Furthermore, increased intragastric pressure was also linked to a higher incidence of staple line leak and GERD, and thus larger bougie size was thought to decrease the risk of leak and GERD [25, 26]. An expert panel consensus [27] concluded that the use of a smaller bougie might be associated with a higher risk of leak, GERD, and other complications. However, a large database study showed a variable impact of the bougie size on the incidence of GERD after SG as larger bougies (≥ 36 F) had a GERD incidence of 28.8–30.7% whereas smaller bougies (< 36 F) had an incidence of 27.5–33.5% [28].

The other technical factor was the distance of the first stapler firing from the pylorus. A meta-analysis [29] found no significant association between the distance from the pylorus and incidence of GERD after SG. However, one main limitation of this meta-analysis was the substantial statistical heterogeneity of the studies included. Interestingly, as shown in this meta-analysis, the odds of developing GERD increased from 1.66 for a distance of 6 cm to 4.79 for a distance of 2–3 cm. This may also be explained on the basis of the intraluminal pressure theory. When the first stapler firing is closer to the pylorus, this would create a narrower gastric tube than if it was farther from the pylorus. Again, with a narrower gastric tube, the intraluminal pressure would be higher and thus increases the risk of GERD.

To expand on the clinical utility of the AI model we developed, we opted to develop a multitask platform that would help in decision-making. This platform identified the cutoff values for age, weight, bougie size, and distance from the pylorus which may help surgeons tailor the modifiable technical factors to reduce the risk of GERD after SG in patients with non-modifiable risk factors as older age and heavier weight.

The limitations of the present study include its retrospective and single-center nature that would be associated with an inherent risk of selection bias and lack of external validity of the results. Larger, multicenter prospective trials are needed to verify the findings of our study and to externally validate the AI model that was developed.

Conclusion

Using patient-related and technical parameters, an AI-based model for the prediction of GERD after SG was developed. The model had excellent accuracy with an AUC of 0.93, yet

a moderate sensitivity of 79.2% and specificity of 86.1%. The top-ranked parameters of the AI model were age, baseline weight, preoperative GERD, size of the orogastric tube, and distance of first stapler firing from the pylorus.

Declarations

Ethics Approval Ethics approval was obtained from the Institutional Review Board of our institution.

Consent to Participation Informed consent was not required owing to the retrospective nature of the study.

Conflict of Interest The authors declare no competing interests.

References

1. Ozsoy Z, Demir E. Which bariatric procedure is the most popular in the world? A Bibliometric Comparison. Obes Surg. 2018;28:2339–52. https://doi.org/10.1007/s11695-018-3163-6.

2. Emile SH, Elfeki H, Elalfy K, et al. laparoscopic sleeve gastrectomy then and now: an updated systematic review of the progress and short-term outcomes over the last 5 years. Surg Laparosc Endosc Percutan Tech. 2017;27(5):307–17. https://doi.org/10.1097/SLE.0000000000000418.

3. Yeung KTD, Penney N, Ashrafian L, et al. does sleeve gastrectomy expose the distal esophagus to severe reflux?: A systematic review and meta-analysis. Ann Surg. 2020;271(2):257–65. https://doi.org/10.1097/SLA.0000000000003275.

4. Stenard F, Iannelli A. Laparoscopic sleeve gastrectomy and gastroesophageal reflux. World J Gastroenterol. 2015;21(36):10348–57. https://doi.org/10.3748/wjg.v21.i36.10348.

5. Sharma A, Aggarwal S, Ahuja V, et al. Evaluation of gastroesophageal reflux before and after sleeve gastrectomy using symptom scoring, scintigraphy, and endoscopy. Surg Obes Relat Dis. 2014;10:600–5.

6. Rebecchi F, Allaix ME, Giaccone C, et al. Gastroesophageal reflux disease and laparoscopic sleeve gastrectomy: a physiopathologic evaluation. Ann Surg. 2014;260:909–14; discussion 914-5.

7. Morshed MM, Abdel-Razik MA, Mahmoud SA, et al. Functional assessment of the esophagogastric junction after laparoscopic sleeve gastrectomy. J Surg. 2016;4(5):105–13. https://doi.org/10.11648/j.js.20160405.14.

8. Emile SH. Gastroesophageal reflux disease after sleeve gastrectomy: the need to predict its onset and prevent its consequences. Obes Surg. 2019;29(8):2625–6. https://doi.org/10.1007/s11695-019-03955-9.

9. Althuwaini S, Bamehriz F, Aldohayan A, et al. Prevalence and predictors of gastroesophageal reflux disease after laparoscopic sleeve gastrectomy. Obes Surg. 2018;28(4):916–22. https://doi.org/10.1007/s11695-017-2971-4.

10. Jonasson C, Wernersson B, Hoff DA, et al. Validation of the GerdQ questionnaire for the diagnosis of gastro-oesophageal reflux disease. Aliment Pharmacol Ther. 2013;37(5):564–72. https://doi.org/10.1111/apt.12204.

11. Gu L, Chen B, Du N, et al. Relationship between bariatric surgery and gastroesophageal reflux disease: a systematic review and meta-analysis. Obes Surg. 2019;29(12):4105–13. https://doi.org/10.1007/s11695-019-04218-3.

12. Wright FG, Duro A, Medici JR, et al. Esophageal adenocarcinoma five years after laparoscopic sleeve gastrectomy. A case report. Int J Surg Case Rep. 2017;32:47–50. https://doi.org/10.1016/j.ijscr.2017.01.054.

13. Bellorin O, Senturk JC, Cruz MV, et al. Predictive factors for developing GERD after sleeve gastrectomy: is preoperative endoscopy necessary? J Gastrointest Surg. 2022. https://doi.org/10.1007/s11605-021-05207-7.

14. Navarini D, Madalosso CAS, Tognon AP, et al. Predictive factors of gastroesophageal reflux disease in bariatric surgery: a controlled trial comparing sleeve gastrectomy with gastric bypass. Obes Surg. 2020;30:1360–7. https://doi.org/10.1007/s11695-019-04286-5.

15. Althuwaini S, Bamehriz F, Aldohayan A, et al. Prevalence and predictors of gastroesophageal reflux disease after laparoscopic sleeve gastrectomy. Obes Surg. 2018;28(4):916–22. https://doi.org/10.1007/s11695-017-2971-4.

16. Steyerberg EW, Eijkemans MJ, Habbema JD. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol. 1999;52:935–42.

17. Grant SW, Hickey GL, Head SJ. Statistical primer: multivariable regression considerations and pitfalls. Eur J Cardiothorac Surg. 2019;55(2):179–85. https://doi.org/10.1093/ejcts/ezy403.

18. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–8. https://doi.org/10.1038/nature21056.

19. Soguero-Ruiz C, Hindberg K, Mora-Jiménez I, et al. Predicting colorectal surgical complications using heterogeneous clinical data and kernel methods. J Biomed Inform. 2016;61:87–96. https://doi.org/10.1016/j.jbi.2016.03.008.

20. Stam WT, Goedknegt LK, Ingwersen EW, Schoonmade LJ, Bruns ERJ, Daams F. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: a systematic review. Surgery. 2021:S0039-6060(21)00960-0.https://doi.org/10.1016/j.surg.2021.10.002

21. Ghareeb WM, Emile SH, Elshobaky A. Artificial intelligence compared to alvarado scoring system alone or combined with ultrasound criteria in the diagnosis of acute appendicitis. J Gastrointest Surg. 2021. https://doi.org/10.1007/s11605-021-05147-2.

22. Soliman H, Coupaye M, Cohen-Sors B, et al. Do preoperative esophageal ph monitoring and high-resolution manometry predict symptoms of GERD after sleeve gastrectomy? Obes Surg. 2021;31(8):3490–7. https://doi.org/10.1007/s11695-021-05427-5.

23. Yamasaki T, Hemond C, Eisa M, et al. The changing epidemiology of gastroesophageal reflux disease: are patients getting younger? J Neurogastroenterol Motil. 2018;24(4):559–69. https://doi.org/10.5056/jnm18140.

24. de Vries DR, van Herwaarden MA, Smout AJ, et al. Gastroesophageal pressure gradients in gastroesophageal reflux disease: relations with hiatal hernia, body mass index, and esophageal acid exposure. Am J Gastroenterol. 2008;103:1349–54.

25. Aurora AR, Khaitan L, Saber AA. Sleeve gastrectomy and the risk of leak: a systematic analysis of 4,888 patients. Surg Endosc. 2012;26(6):1509–15.

26. Rebecchi F, Allaix ME, Patti MG, et al. Gastroesophageal reflux disease and morbid obesity: to sleeve or not to sleeve? World J Gastroenterol. 2017;23(13):2269–75. https://doi.org/10.3748/wjg.v23.i13.2269.

27. Rosenthal RJ. International sleeve gastrectomy expert panel, international sleeve gastrectomy expert panel consensus statement: best practice guidelines based on experience of >12,000 cases. Surg Obes Relat Dis. 2012;8(1):8–19. https://doi.org/10.1016/j.soard.2011.10.019.

28. Haskins IN, Jackson HT, Graham AE, et al. The effect of bougie size and distance from the pylorus on dehydration after laparoscopic sleeve gastrectomy: an analysis of the ACS-MBSAQIP database. Surg Obes Relat Dis. 2019;15(10):1656–61. https://doi.org/10.1016/j.soard.2019.08.014.

29. Yang J, Huang W, Chen C, et al. Gastroesophageal reflux disease and sleeves gastrectomy with different distance to the pyloric: a systematic review and meta-analysis. Clin Surg. 2020;5:2818.

Ⓐ Springer