Predicting Response to Neuromodulators or Prokinetics in Patients With Suspected Gastroparesis Using Machine Learning: The "BMI, Infectious Prodrome, Delayed GES, and No Diabetes" Model

Will Takakura, MD[1], Brian Surjanhata, MD[2], Linda Anh Bui Nguyen, MD, FACG[3], Henry P. Parkman, MD, FACG[4], Satish S.C. Rao, MD, PhD, FACG[5], Richard W. McCallum, MD, FACG[6], Michael Schulman, DO, FACG[7], John Man-Ho Wo, MD[8], Irene Sarosiek, MD, FACG[6], Baha Moshiree, MD, FACG[9], Braden Kuo, MD, FACG[2], William L. Hasler, MD[10] and Allen A. Lee, MD, MS[1]

INTRODUCTION: **Pharmacologic therapies for symptoms of gastroparesis (GP) have limited efficacy, and it is difficult to predict which patients will respond. In this study, we implemented a machine learning model to predict the response to prokinetics and/or neuromodulators in patients with GP-like symptoms.**

METHODS: **Subjects with suspected GP underwent simultaneous gastric emptying scintigraphy (GES) and wireless motility capsule and were followed for 6 months. Subjects were included if they were started on neuromodulators and/or prokinetics. Subjects were considered responders if their GP Cardinal Symptom Index at 6 months decreased by ≥1 from baseline. A machine learning model was trained using lasso regression, ridge regression, or random forest. Five-fold cross-validation was used to train the models, and the area under the receiver operator characteristic curve (AUC-ROC) was calculated using the test set.**

RESULTS: **Of the 150 patients enrolled, 123 patients received either a prokinetic and/or a neuromodulator. Of the 123, 45 were considered responders and 78 were nonresponders. A ridge regression model with the variables, such as body mass index, infectious prodrome, delayed gastric emptying scintigraphy, no diabetes, had the highest AUC-ROC of 0.72. The model performed well for subjects on prokinetics without neuromodulators (AUC-ROC of 0.83) but poorly for those on neuromodulators without prokinetics. A separate model with gastric emptying time, duodenal motility index, no diabetes, and functional dyspepsia performed better (AUC-ROC of 0.75).**

DISCUSSION: **This machine learning model has an acceptable accuracy in predicting those who will respond to neuromodulators and/or prokinetics. If validated, our model provides valuable data in predicting treatment outcomes in patients with GP-like symptoms.**

KEYWORDS: predictive model; gastroparesis; prokinetics; neuromodulators; gastric emptying

SUPPLEMENTARY MATERIAL accompanies this paper at http://links.lww.com/CTG/B166, http://links.lww.com/CTG/B167

Clinical and Translational Gastroenterology 2024;15:e00743. https://doi.org/10.14309/ctg.0000000000000743

INTRODUCTION

Gastroparesis (GP) is a chronic condition characterized by delayed gastric emptying in the setting of symptoms, including nausea, vomiting, bloating, fullness, early satiety, and abdominal pain. Despite the morbidity (1) associated with GP, treatment options are limited with only metoclopramide being approved by

[1]Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan, USA; [2]Division of Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts, USA; [3]Division of Gastroenterology, Stanford Medicine, Stanford, California, USA; [4]Division of Gastroenterology, Temple University Health System Inc, Philadelphia, Pennsylvania, USA; [5]Division of Gastroenterology, Augusta University, Augusta, Georgia, USA; [6]Division of Gastroenterology, Texas Tech University Health Sciences Center El Paso, El Paso, Texas, USA; [7]Suncoast GI Associates LLC, Bradenton, Florida, USA; [8]Division of Gastroenterology, Indiana University School of Medicine, Indianapolis, Indiana, USA; [9]Gastroenterology and Hepatology, Atrium Health, Charlotte, North Carolina, USA; [10]Division of Gastroenterology, Mayo Clinic, Scottsdale, Arizona, USA. **Correspondence:** Allen A. Lee, MD, MS. E-mail: allenlee@med.umich.edu.

Received June 7, 2024; accepted June 14, 2024; published online July 1, 2024

the US Food and Drug Administration (2). A recent network meta-analysis concluded that only clebopride and domperidone were superior to placebo (3). Both medications are not approved by the US Food and Drug Administration. Treatment with prokinetic agents and/or neuromodulators based on physiologic testing (4), but the evidence supporting this practice is not robust, and neuromodulators are not recommended for GP treatment in guidelines (5,6). This leads to patients being trialed on numerous ineffective medications, while patients harbor the cost and burden of ineffective therapy.

Identifying patients more likely to respond to therapies in suspected GP has been challenging. The correlation between improvement in delayed gastric emptying and symptomatic improvement has been relatively weak (7), and correlation between symptoms and delayed gastric emptying has been modest at best (8). In addition, most medications that are associated with both improvements in symptoms and gastric emptying are not readily available in the United States (9). A prior study by the National Institute of Health GP Consortium identified age 50 years or older, gastric emptying scintigraphy (GES) retention of ≥20% at 4 hours, and infectious prodrome as predictors of improved outcomes in GP (10). Our group also recently published predictors of longitudinal outcomes in suspected GP, including female gender, delayed gastric emptying, presence of functional dyspepsia, and harder stools as predicting worse outcomes in suspected GP (11). Although these studies improved our knowledge of longitudinal outcomes in GP, they did not identify predictors of response to therapies.

Recently, there has been a growing interest in applying machine learning models to predict clinical outcomes in medicine. In stroke, a machine learning model accurately predicted long-term outcomes with an area under the receiver operator characteristic curve (AUC) of 0.89 (12). In pediatric critical care, acute kidney injury was projected 30 hours before conventional detection with an AUC of 0.89 (13). A deep-learning model accurately predicted cirrhosis based on electrocardiogram with an AUC of 0.91 (14). Given the poor response to therapy in GP, we hypothesized that machine-learning models may identify subsets of patients more likely to respond to different therapies in GP.

In this multicenter prospective study, we aimed to develop a predictive model to identify responses to neuromodulators and/or prokinetics in patients with GP-like symptoms.

METHODS

Study population

We performed a prospective, observational cohort study of 150 adult subjects with 2 or more gastroparetic symptoms (nausea/vomiting/retching, fullness/early satiety, bloating/abdominal distention, upper abdominal discomfort/pain) for 12 or more weeks. Subjects were recruited from 2013 to 2016 at 10 academic and community centers in the United States (15). Subjects underwent simultaneous GES and wireless motility capsule (WMC) test at baseline and were given treatment recommendations at the discretion of the provider (4). Subjects were followed, and their symptoms were assessed later at 6 months.

We have reported portions of this study (ClinicalTrials.gov: NCT02022826), including validation of WMC (15), the influence of motility test results on management decisions (4), longitudinal outcomes (11), and the effects of prokinetics (16). The aim of this study was to identify predictors of response to different medication classes, which were a priori planned study end points. Of the initial 150 subjects, only those who received a neuromodulator and/or a prokinetic were included in this analysis. Concomitant use of other medications such as laxatives, antiemetics, and dietary therapies were allowed based on the discretion of the prescribing providers.

Primary outcomes

Our primary outcome of interest was response to neuromodulators and/or prokinetics as defined by a reduction in the Gastroparesis Cardinal Symptom Index (GCSI) score by ≥1 at 6 months compared with baseline. GCSI is a validated questionnaire comprised 9 questions to assess the gastroparetic symptoms of nausea/vomiting, fullness/early satiety, and bloating/distention from 0 to 5, with 5 being the worst symptoms, in patients with and without delayed gastric emptying (17). We chose our definition for response based on a prior study in GP which determined that the minimal clinically important difference in GCSI scores to be 0.94 (18). Subjects who did not complete GCSI scores at 6 months were considered to be nonresponders.

Motility testing. GES and WMC were performed simultaneously at baseline. Subjects were asked to hold their proton-pump inhibitors for 7 days while histamine-2 antagonists, prokinetics, opioids, cannabinoids, and laxatives were held for 3 days before testing. After fasting overnight, subjects consumed ^{99}Tc-radiolabeled standardized low-fat egg substitute meals (19). Immediately after, subjects swallowed the WMC (SmartPill; Medtronic; Minneapolis, MN) with 50 mL of water and followed the manufacturer's instructions. Anterior and posterior scintigraphy images were obtained at 0, 1, 2, and 4 hours after ingestion of the test meal. Subjects were instructed not to eat for 8 hours after capsule ingestion, followed by ingestion of 250 mL of a liquid nutrient drink (Ensure; Abbott Laboratories; Abbott Park, IL). Subjects fasted an additional hour before resuming their diet. WMC receiver was worn for up to 5 days or until the capsule was seen in the toilet.

Predictor variables. Baseline characteristics that were evaluated as potential predictors included age, gender, body mass index (BMI), history of diabetes, cannabis use, opioid use, duration of GP symptoms, stool consistency as measured by Bristol Stool Scale, and history of infectious prodrome (per patient interview). In addition, Rome III criteria for functional bowel disorders, functional dyspepsia, functional nausea/vomiting/belching disorders, and constipation were evaluated as potential predictors in the model. Concomitant treatments, such as a gastroparetic diet, antiemetic, and laxative use, were also captured. Finally, motility parameters and other specific features obtained from GES and WMC were calculated and interpreted for further analysis. Delayed GES was defined as >10% retention at 4 hours (19), and delayed gastric emptying time (GET) through WMC was defined as >5 hours after ingestion of the capsule (20). The number of contractions and motility index (MI) by WMC were measured in the hour before and after GET to determine stomach and small bowel contractile parameters (21).

Data preprocessing. Data were randomly selected with 75% of the patients used for model training, and the remaining 25% of the patients held out for testing and validation of model performance. The data were stratified by the proportion of responders so that the distribution of the outcome was similar in both the training

and the test set. Missing data were imputed using the R package *missForest*, a random forest-based multiple imputation method previously shown to have the lowest imputation error for both continuous and categorical variables (22). Numeric variables were centered and scaled while categorical variables were recoded into dummy variables. To select predictors to include in our model, we performed a univariable analysis measuring the association between each predictor and our primary outcome, reduction in GCSI of ≥1 at 6 months compared with baseline. Those with $P < 0.25$ were included as potential features in the model selection process. The number of predictors was further reduced using the step_select_linear function in the R package *recipeselectors* (23).

Model development and testing. Using a *Tidymodels* framework (24), 3 separate machine learning classification models were trained using the randomly selected 75% of subjects included in the training data. We performed penalized regression, including least absolute shrinkage and selection operator (Lasso) and ridge regression, with the R package *glmnet* (25). We also fit a random forest model with the R package *ranger* (26). Machine learning

Table 1. Baseline variables

	Neuromodulators and/or prokinetics (n = 123)	Neuromodulators (n = 50)	Prokinetics (n = 52)
Baseline variables			
Age	44 (34–54)	41.5 (31.8–51.0)	47.0 (37.0–56.3)
Female	98 (79.7%)	40 (80.0%)	44 (84.6%)
BMI	26.6 (22.2–31.5)	25.6 (21.2–31.4)	27.8 (23.0–31.3)
Marijuana	10 (8.1%)	2 (4.0%)	6 (11.5%)
Opioids	14 (11.4%)	7 (14.0%)	4 (7.7%)
Laxatives	44 (35.8%)	17 (34.0%)	15 (28.9%)
Antiemetics	15 (12.2%)	3 (6.0%)	9 (17.3%)
Gastroparesis diet	44 (35.8%)	12 (24.0%)	26 (50.0%)
Delayed GES	32 (26.0%)	4 (8%)	22 (42.3%)
Delayed GET	52 (42.3%)	6 (12.0%)	35 (67.3%)
GES % retention at 4 hr	3 (1–11)	1 (1–4)	5.1 (2.8–21.8)
GES % retention at 2 hr	33 (18.2–52.5)	26.5 (13.3–37.8)	41.5 (27.8–58.3)
GET	4.4 (3.3–6.7)	3.5 (2.9–4.4)	6.5 (4.4–18.5)
Small bowel transit time (hr)	4.71 (3.54–5.87)	4.5 (4.0–5.4)	4.9 (3.5–6.5)
Colonic transit time (hr)	37.69 (17.1–68.0)	30.0 (17.1–55.9)	43.5 (19.4–70.0)
No. of antral contractions	45.5 (24.0–87.3)	50.0 (26.5–88.5)	46.0 (29.0–93.8)
Antral motility index	11.3 (10.1–12.6)	11.3 (10.1–12.7)	11.4 (10.2–12.7)
No. of duodenal contractions	104 (46–166)	126.5 (58.5–175.0)	86.0 (42.5–159.5)
Duodenal motility index	12.4 (11.0–13.6)	12.8 (11.1–13.7)	12.3 (10.5–13.6)
Infectious prodrome	15 (12.2%)	7 (14.0%)	5 (9.6%)
Duration of GP symptoms (mo)	36 (21–84)	36 (24–72)	36 (24–84)
Diabetes	33 (26.8%)	10 (20.0%)	34 (65.4%)
Bristol stool form scale	4 (2–6)	4 (2–6)	4 (2–5.3)
Functional bowel disorder	96 (78.1%)	39 (78.0%)	41 (78.9%)
Constipation	40 (32.5%)	15 (30%)	21 (40.4%)
Functional dyspepsia	89 (72.4%)	35 (70.0%)	39 (75%)
Nausea vomiting and belching	75 (61.0%)	31 (62.0%)	32 (61.5%)
Baseline GCSI	2.81 (2.17–3.23)	2.81 (2.23–3.1)	2.79 (2.13–3.23)
Nausea/vomiting subscore	1.33 (0.67–2.83)	1.33 (0.67–2.00)	1.5 (0.67–3.00)
Bloating/distention subscore	3.50 (2.25–4.00)	3.50 (3.00–4.00)	3.00 (2.00–4.00)
Fullness/satiety subscore	3.25 (2.75–4.00)	3.25 (2.50–4.00)	3.50 (2.75–4.00)
Upper abdominal pain subscore	3.00 (2.00–4.00)	3.00 (1.65–4.00)	3.00 (1.50–4.00)

Baseline variables for those who received neuromodulators or prokinetics, neuromodulators without prokinetics, and prokinetics without neuromodulators. Continuous variables are expressed as median (interquartile range).
BMI, body mass index; GCSI, Gastroparesis Cardinal Symptom Index; GES, gastric emptying scintigraphy; GET, gastric emptying time; GP, gastroparesis.

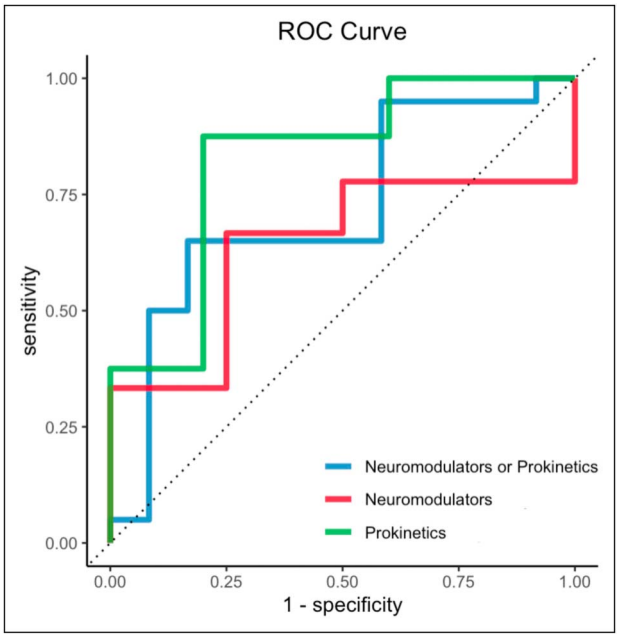

Figure 1. AUC-ROC of the final model to predict a response to prokinetics or neuromodulators with the predictors delayed GCSI, diabetes, infectious prodrome, and BMI using ridge regression. The AUC-ROC were 0.72 for neuromodulators or prokinetics (blue line), 0.64 for neuromodulators without prokinetics (red line), and 0.83 for prokinetics without neuromodulators (green line). AUC-ROC, area under the receiver operator characteristic curve; BMI, body mass index; GCSI, Gastroparesis Cardinal Symptom Index.

algorithms were first applied to the training data to parameterize and fit the model. Five-fold cross-validation was utilized to estimate model accuracy and tune model hyperparameters. Model accuracy was then evaluated by calculating the AUC using the independent test data consisting of the remaining 25% of patients not selected for the training set.

Variable importance. Variable importance from each model was determined by using the R package *vip*, which provides model-specific variable importance scores (27). We also performed locally interpretable model-agnostic explanations using the R package *breakdown* (28), which decomposes model predictions into parts that can be attributed to different explanatory variables.

Sensitivity analysis

To determine whether our model was generalizable, we performed a sensitivity analysis using a different GCSI cutoff ≥0.75 as responders. In addition, a subgroup analysis using the same model was applied to those who received prokinetics without neuromodulators and neuromodulators without prokinetics. Finally, to determine if model performance was dependent on the modality of gastric motility test, GES values were substituted for equivalent WMC parameters.

RESULTS

Of the 150 subjects, 123 subjects were prescribed either a neuromodulator and/or a prokinetic and were included in the analysis. Fifty patients received neuromodulators without prokinetics, 52 received prokinetics without neuromodulators and 21 received both. Of the 123 subjects, 45 subjects were considered responders and 78 were considered nonresponders. Baseline variables for possible incorporation into the model, as well as the GCSI score at baseline and 6 months are described in Table 1. Notably, significantly more subjects in the GCSI responder group were diabetic with 17 (37.8%) vs 16 (20.5%), $P = 0.04$ (see Supplementary Table 1, http://links.lww.com/CTG/B167). At 6 months, those in the responder group had an improvement in their median (interquartile range) GCSI from 2.7 (2.0–3.1) to 1.2 (0.8–1.8) compared with 3.1 (2.5–3.5) to 2.8 (2.0–3.6) in the nonresponder group.

Variable selection

We performed feature selection to avoid multicollinearity and prevent overparameterization as well as to increase clinical utility of the predictive models. The predictors that were associated with the primary outcome ($P < 0.25$) were BMI, presence of infectious prodrome, history of diabetes, delayed GES, and meeting Rome III criteria for nausea, vomiting, and retching. In addition, predictors were ranked according to step_select_linear function, and predictors with the highest coefficients were considered for inclusion into the final model, which were BMI, presence of infectious prodrome, history of diabetes, and delayed GES.

Prediction for prokinetics and/or neuromodulators. A ridge regression model consisting of 4 variables: **B**MI, **i**nfectious prodrome, **d**elayed GES, and **n**o **d**iabetes (or BIDND) had the highest AUC by 5-fold cross-validation. This model showed acceptable accuracy when tested on the independent test set (AUC = 0.72, Figure 1).

A lasso regression model incorporating the BIDND predictors also showed a similar yet lower accuracy when tested on the independent test set (AUC = 0.69, see Supplementary Figure 1, http://links.lww.com/CTG/B166). However, a random forest model fitted for the BIDND predictors performed very poorly (AUC = 0.49). Given that the ridge regression model had the highest performance, this model was carried forward for further analyses.

Predictive features for response to neuromodulators and/or prokinetics. Delayed GES was the most important predictor by variable importance analysis, followed by absence of diabetes, BMI, and infectious prodrome (Figure 2a). Next, we generated breakdown plots to explain the contribution of each feature to the model prediction, which showed that delayed GES, absence of diabetes, and infectious prodrome were predictive of nonresponse, while higher BMI was predictive of response to neuromodulators and/or prokinetics (Figure 2b,c).

Choice of gastric emptying testing does not impact prediction. We next determined whether the modality of gastric emptying testing impacted model results. We found model performance was similar when using GES results as a binary (i.e., delayed vs nondelayed) or continuous outcome (i.e., percent retention at 4 hours) with AUC of 0.72 and 0.77, respectively (Figure 3a). Similarly, model performance was acceptable when delayed GES was replaced with GETs by WMC as a binary or continuous outcome (AUC 0.70 and 0.73, respectively) (Figure 3b).

Sensitivity analysis of BIDND model with various cutoffs for GCSI

Given the differences in definition for responders in the GP literature with 1 study suggesting a minimally clinically significant

difference in GCSI score was 1 (18) while a prior study suggested 0.75 (17), we sought to determine whether model performance was affected by the definition for responder. Using a lower threshold for responder (i.e., change in GCSI > 0.75), there were 52 responders and 71 nonresponders in our cohort. With this lower threshold for the responder, the BIDND model performance remained acceptable (AUC 0.70) (Figure 3c).

Prediction for response to prokinetics without neuromodulators. Given the clinical interest in identifying subjects likely to respond specifically to neuromodulators or prokinetics, we next evaluated how well the BIDND model predicted response in subjects receiving prokinetics but not neuromodulators. Of the 52 subjects who received prokinetics but without neuromodulators, 20 (38.5%) were responders. The BIDND model showed good performance for predicting response to prokinetics without neuromodulators (AUC = 0.83, Figure 1).

Variable importance showed history of infectious prodrome was the most important predictor, followed by delayed GES, no diabetes, and BMI (Figure 4a). Similar to the overall model in neuromodulators and/or prokinetics, absence of infectious prodrome, normal gastric emptying, and absence of diabetes increased likelihood of response to prokinetics (Figure 4b,c).

Prediction for response to neuromodulators without prokinetics. We also evaluated how well the BIDND model performed in predicting response to neuromodulators without prokinetics. Of the 50 subjects who received neuromodulators but without prokinetics, 14 (28.0%) were responders using the GCSI threshold of >1. BIDND model performance in those receiving neuromodulators without prokinetics was poor (AUC = 0.64, Figure 1).

Given the poor predictive ability of the BIDND model in this subgroup and the clinical utility in identifying response specifically to neuromodulators, we explored whether a separate model with unique predictors would show better performance for those subjects prescribed a neuromodulator without prokinetics. We utilized a similar method by first performing a univariable analysis and selecting 3 predictors with the lowest *P* values. Given prior data supporting an association between functional dyspepsia (FD) and response with neuromodulators, we also included presence of functional dyspepsia by Rome III criteria into this new model (29). This model including GET, duodenal MI, absence of diabetes, and functional dyspepsia showed acceptable performance for predicting response to neuromodulators without prokinetics (AUC 0.75, Figure 5a). The absence of diabetes, longer GET, presence of FD were negative predictors while duodenal MI was a predictor of response to neuromodulators without prokinetics (Figure 5b–d).

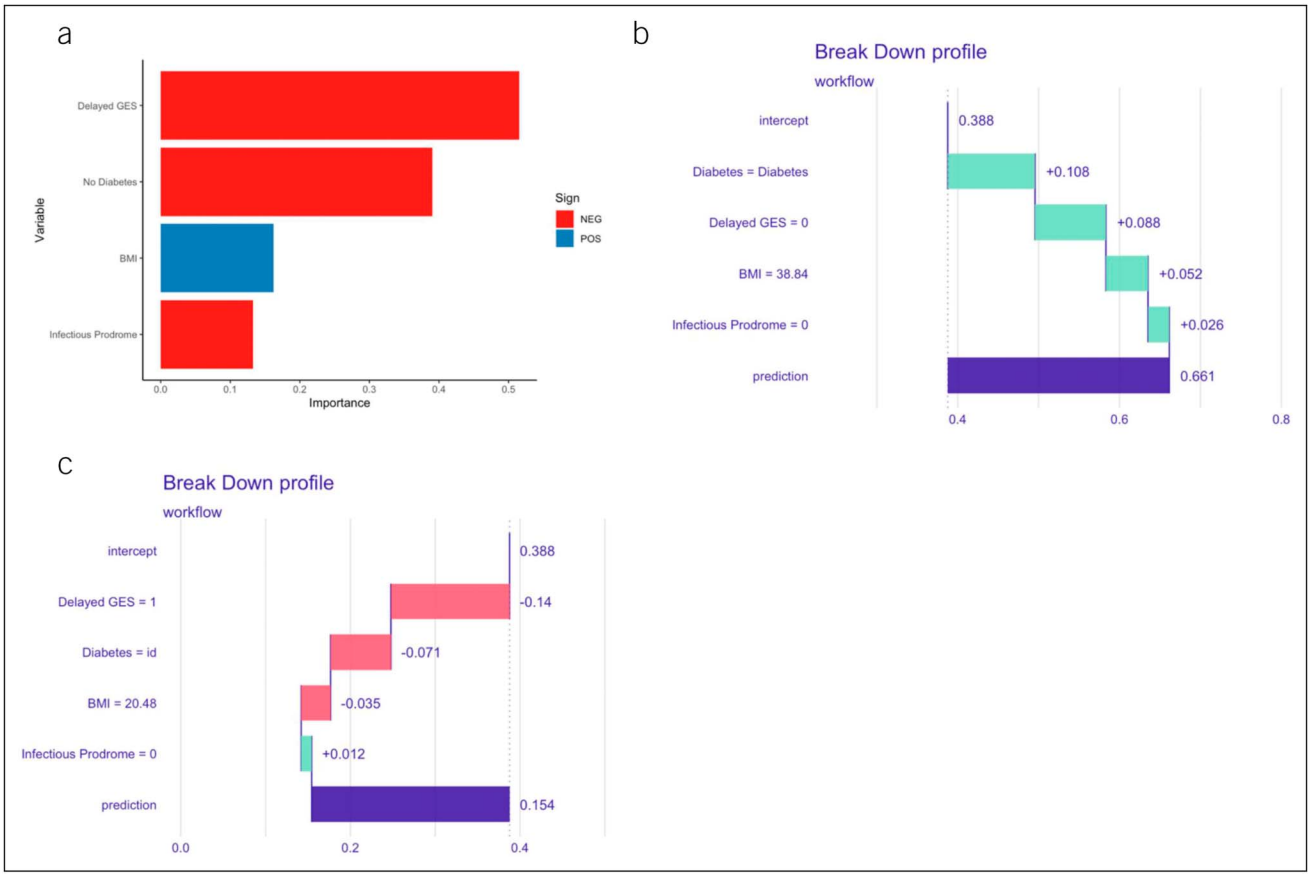

Figure 2. (**a**) VIP for neuromodulator and/or prokinetics. The most important variables were delayed GES, no diabetes, and BMI, followed by infectious prodrome. Break down plots for subjects with (**b**) high and (**c**) low predicted probability for response to neuromodulators and/or prokinetics. The intercept represents the mean model-specific predicted probability for response to neuromodulators and/or prokinetics while each subsequent variable increases (*green* bar) or decreases (*red* bar) the predicted probability and results in the overall predicted probability (*purple* bar, labeled prediction). BMI, body mass index; GES, gastric emptying scintigraphy; id, idiopathic; VIP, variable importance plot.

DISCUSSION

In this prospective multicenter cohort study, we created a predictive model using machine learning algorithms to predict response to prokinetics and/or neuromodulators. Our BIDND model demonstrated good performance when tested on an independent test set (AUC 0.72). Furthermore, the BIDND model showed good performance in predicting response to prokinetics without neuromodulators while model performance was poor when predicting response to neuromodulators without prokinetics. A separate model comprising GET, diabetes, duodenal MI, and FD had an acceptable AUC for response to neuromodulators without prokinetics. Delayed GES, absence of diabetes, and infectious prodrome were predictors of nonresponse, while increase in BMI was predictive of a response. Notably, our BIDND model had similar accuracy when interchanging GES with GET and worked well when applied to those receiving prokinetics without neuromodulators.

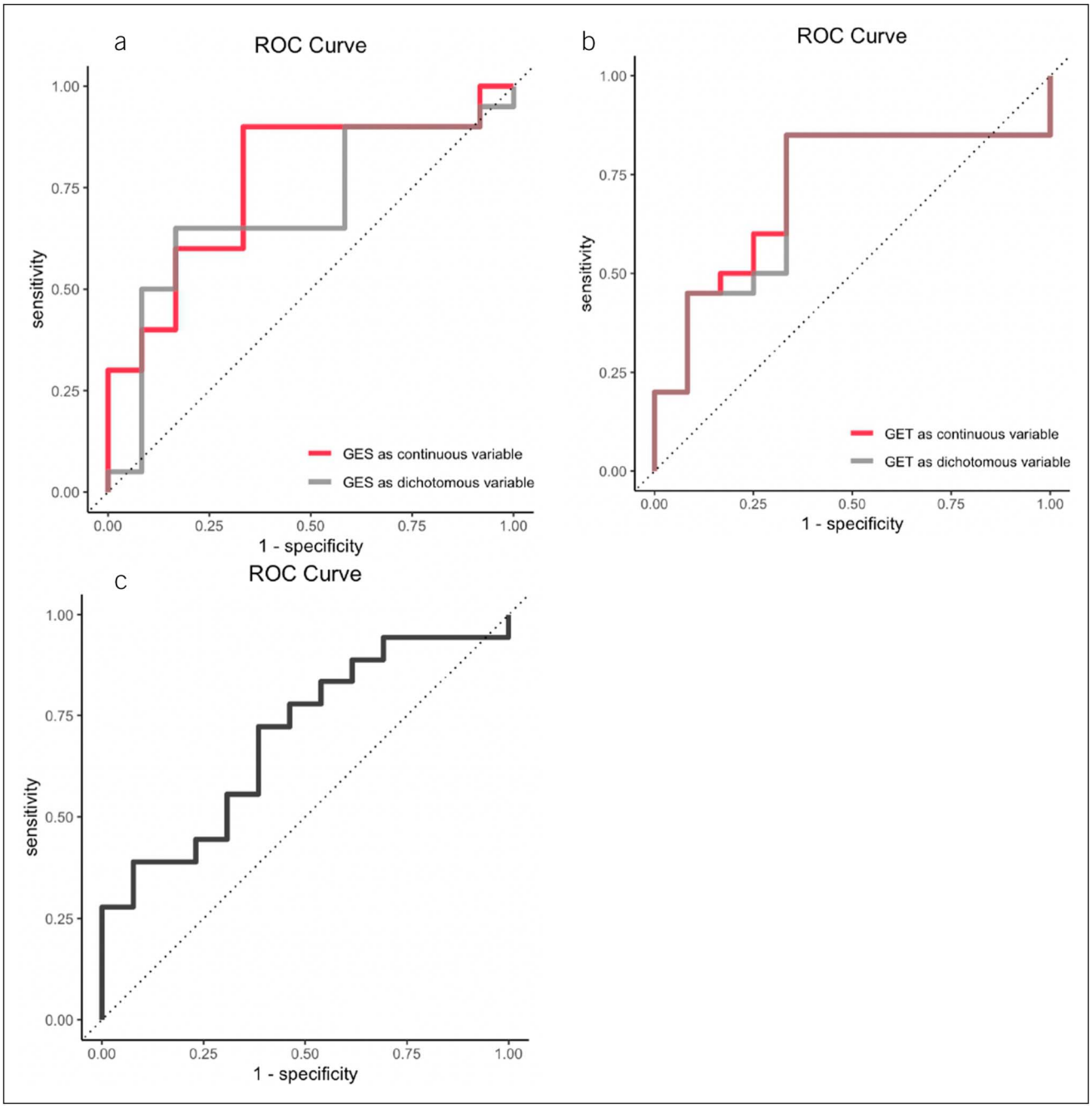

Figure 3. (**a**) AUC-ROC using GES as dichotomous variable (delayed or not delayed) was 0.72 (gray line) and 0.77 using GES as a continuous variable (% retention at 4 hour) (red line). (**b**) AUC-ROC substituting GES for GET through WMC was similar with AUC of 0.7 as dichotomous variable (delayed vs not delayed) and 0.73 as continuous variable. (**c**) AUC-ROC with a lower threshold of a change in GCSI ≥0.75 was 0.7. AUC-ROC, area under the receiver operator characteristic curve; GCSI, Gastroparesis Cardinal Symptom Index; GES, gastric emptying scintigraphy; GET, gastric emptying time; WMC, wireless motility capsule.

Although there is a clear need, there are few studies that have identified predictors of response to treatment options for patients with suspected GP. One previous retrospective study by Anaparthy et al. showed that nausea, distention, and GCSI score at baseline were associated with response using a logistic regression model (30). Delayed GES were not associated with response in this study. A second prospective, multicenter study conducted by Pasricha et al. (10) showed that a model incorporating male gender, age 50 years and older, overall GCSI score, GES retention of ≥20%, and infectious prodrome were predictors of response at 48 weeks while BMI ≥25, moderate/severe abdominal pain, and smoking were associated with nonresponse. However, there are significant differences between these 2 prior studies and our current study. First, while Anaparthy and Pasricha et al. utilized logistic regression, we utilized ridge regression, a form of machine learning that imposes a penalty to avoid overfitting and has been shown to outperform traditional logistic regression in complex disease (25,31,32). Second, we specifically did not incorporate baseline GCSI or any of its components as predictors in our model. Doing so would potentially introduce data leakage to the model, in which information about the outcome is inadvertently used to train the model and subsequently results in poor predictive performance when tested on new populations (33). Third, unlike the study by Anaparthy et al., we performed a prospective, longitudinal cohort study recruiting from 10 academic and community centers across the United States Furthermore, in contrast to the study by Pasricha et al. where most patients had delayed gastric emptying, only about 1/3rd of patients in our cohort had delayed gastric emptying. Thus, our results can be applied more broadly to patients who present with suspected GP. In addition, our study evaluated predictors to prokinetics and/or neuromodulators, which likely has more clinical translation compared with prediction of overall outcomes. Finally and perhaps most importantly, our study design utilized training data to train the algorithm and independent test data to measure model accuracy. Although Pasricha et al. employed cross-validation to estimate performance of their predictive model, the lack of an independent validation data set may have over-estimated their model performance (34).

In addition, several studies have evaluated factors that are associated with worse outcomes in GP, which is similar to the data presented in this study. Delayed GES and delayed GET were associated with worse GCSI scores at follow-up, which was published from our cohort previously (11). Another retrospective study has shown that delayed GES was associated with a lack of improvement at 4 weeks of follow-up (35). In addition, a prior study in functional dyspepsia demonstrated that response to amitriptyline was increased in patients with normal gastric emptying (29) These results are consistent with our model which showed that delayed GES and/or delayed GET were predictive of a nonresponse. Delayed GES and/or delayed GET may be a tool to predict nonresponse and overall poor prognosis with patients who present with suspected GP. In practice, ordering gastric emptying testing may help physicians prognosticate and predict the response to neuromodulators or prokinetics.

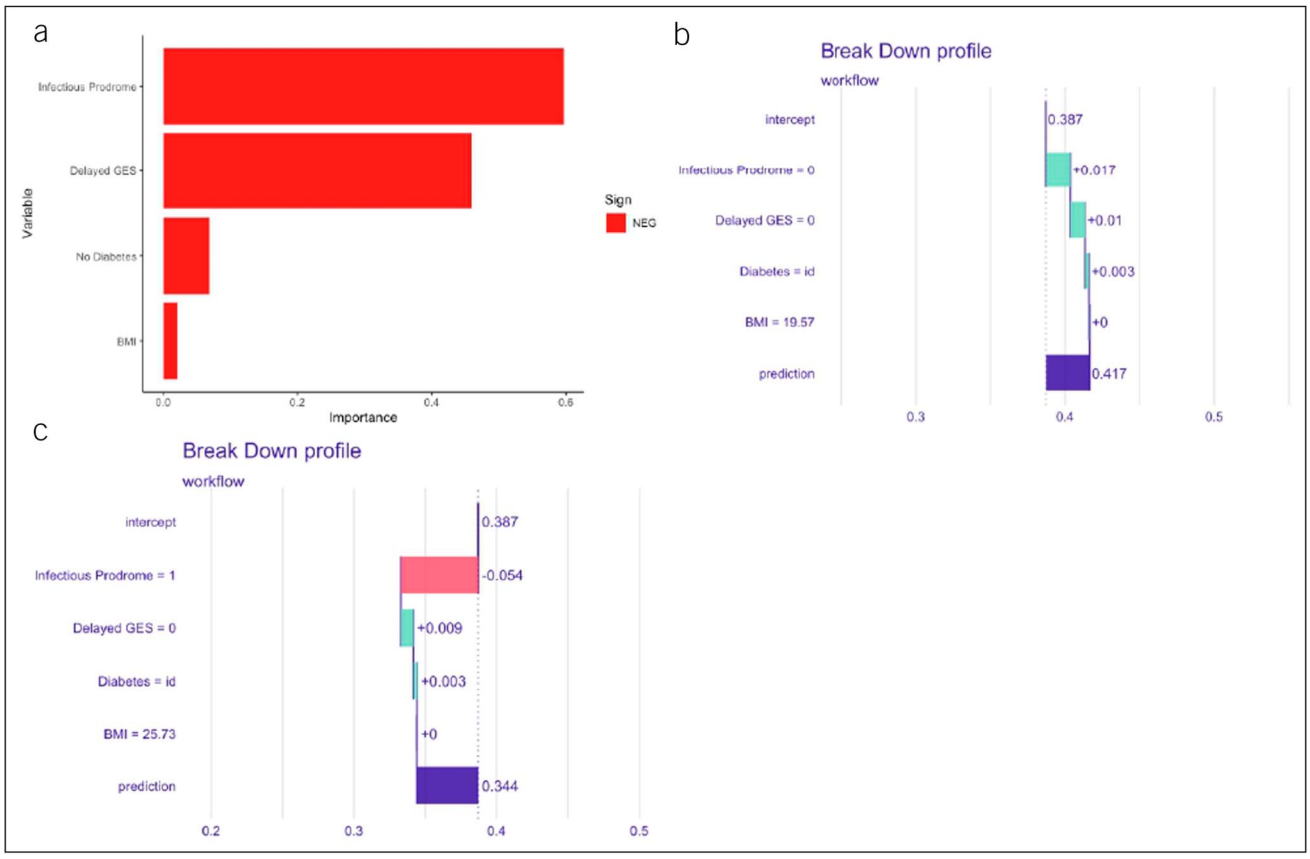

Figure 4. (**a**) VIP plot for prokinetics without neuromodulators. The most important variables were infectious prodrome, diabetes, delayed GES, and followed by BMI. Breakdown plot for subjects with (**b**) high and (**c**) low predicted probability for response to prokinetics without neuromodulators. BMI, body mass index; GES, gastric emptying scintigraphy; id, idiopathic; VIP, variable importance plot.

Interestingly, lack of infectious prodrome was associated with increased likelihood of responding to neuromodulators and/or prokinetics. This contrasts with prior reports demonstrating an infectious prodrome was associated with improved outcomes (36) and better prognosis overall (10). However, a prior study showed that in those GP patients with acute onset symptoms, of which 27.1% had an infectious prodrome, the vast majority (86.9%) continued to have at least moderate-severe symptoms (37). Thus, while many patients with post-infection GP show clinical improvement, in those patients who continue to experience on-going symptoms, an infectious prodrome may be a negative predictor of response to neuromodulators and/or prokinetics. Similarly, in our model comprising GET, absence of diabetes, duodenal MI and FD, and presence of FD decreased the likelihood of response to neuromodulators without prokinetics. Nortriptyline did not improve symptoms in GP (38), while amitriptyline showed benefit in FD, particularly in those with predominant symptoms of abdominal pain (29). Although we did not have data on subtypes of FD in this study, we speculate that patients in our cohort were more likely to have postprandial distress syndrome. Although not entirely similar, this subset of patient may share a similar phenotype to the dysmotility, a subtype of FD that was less likely to respond to amitriptyline (29).

Another novel finding is that an increase in duodenal MI was important in predicting a response to neuromodulator. While the predictive ability of small bowel contractile parameters is unknown, prior studies have shown patients with GP have a blunted duodenal MI after meal ingestion (39) suggestive of neuropathic changes while duodenal contractility measured by WMC was negatively correlated with symptom severity (40). Thus, we speculate that increased duodenal MI may be a favorable prognostic factor and may predict improved response to neuromodulators.

There are several strengths of our study, including the prospective, longitudinal cohort study design utilizing validated outcome measures, including GCSI scores. In addition, our recruitment from multiple academic and community centers across the United States allowed for greater generalizability of our results. Second, we followed best practices for predictive modeling, including utilizing cross-validation for training and developing a model, followed by validating model accuracy using an independent test set. Third, while explaining predictions from machine learning models is difficult, we utilized global and local methods to understand which features were important to the model and the directionality of these features. Finally, we demonstrated that the model performance remained robust when

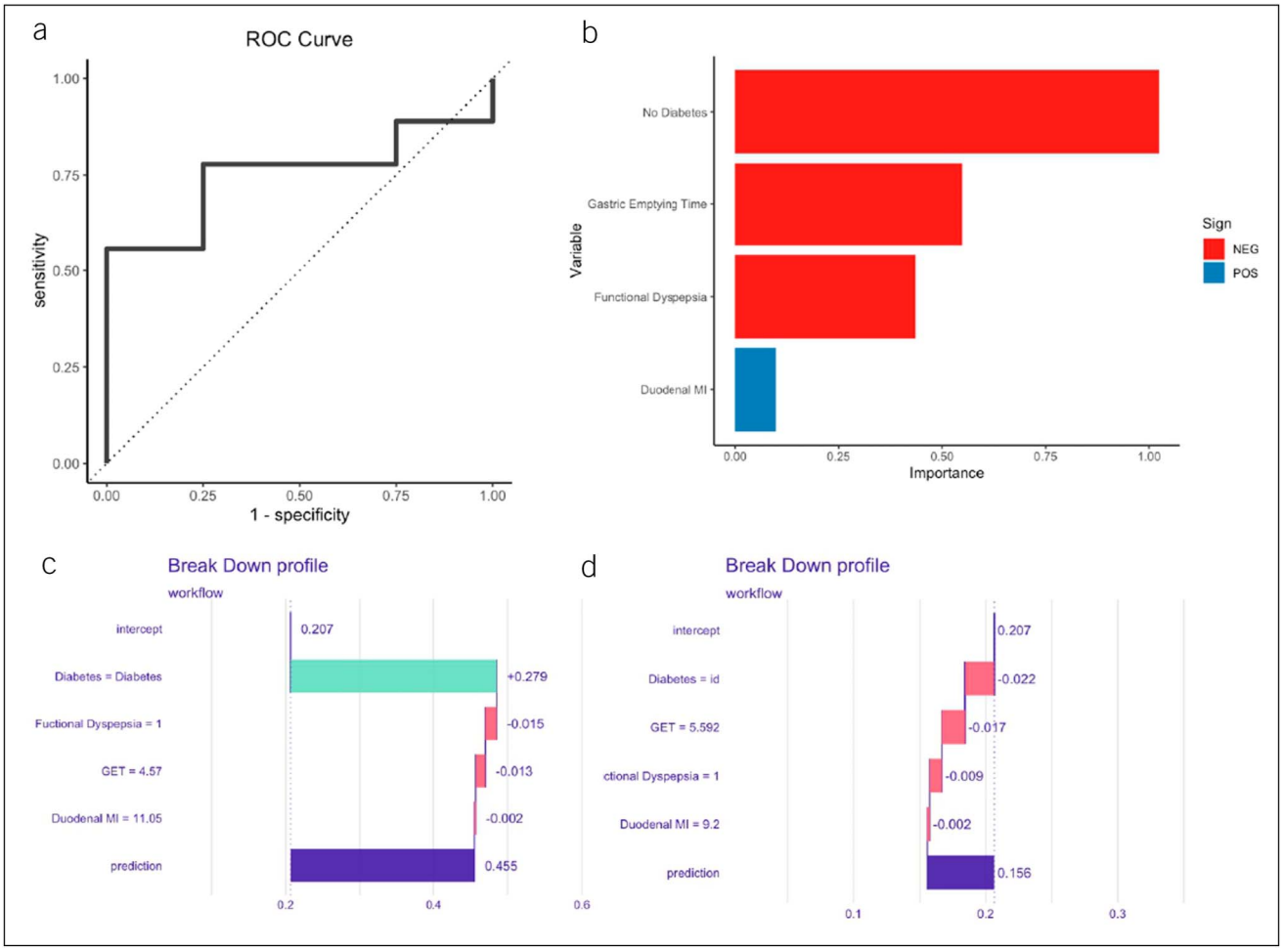

Figure 5. (**a**) ROC curve with predictor variables derived from those on neuromodulator without prokinetics. (**b**) VIP of the model. Breakdown plot for subjects with (**c**) high and (**d**) low predicted probability for response to neuromodulators without prokinetics. ROC, receiver operator characteristic curve; GET, gastric emptying time; id, idiopathic; MI, motility index; VIP, variable importance plot.

interchanging GES and GET and using a different GCSI threshold for response (0.75 vs 1).

However, our study had limitations. Our design allowed the use of other medications such as laxatives, antiemetics, and gastroparetic diets. This limits the interpretation of our model for those who are prescribed exclusively neuromodulators or prokinetics. However, in practice, we often prescribe multiple therapies to fit the need of our patient, and therefore, our model gives real-world predictions. In addition, as medications were selected based on physician preference, our model should not replace physician decision making. Instead, our model may help to predict those who will respond to the selected therapy. Finally, a more restricted analysis of those prescribed only neuromodulators or prokinetics was not possible due to the small sample size. As such, this model should be externally validated with a larger sample set before it can be implemented clinically.

In conclusion, a predictive model with 4 variables BIDND had acceptable accuracy for predicting response to neuromodulators and/or prokinetics in subjects with suspected GP. However, this BIDND model needs to be externally validated using a large multicenter cohort. If validated, this tool may be a valuable resource for clinicians to predict the response of prokinetics and/or neuromodulators in patients with suspected GP.

CONFLICTS OF INTEREST

Guarantor of the article: Allen A. Lee, MD, MS.

Specific author contributions: B.S., L.A.A.N., H.P.P., S.S.C.R., R.W.M., M.S., J.M.W., I.S., B.M., B.K., W.H., A.L.: planning and/or conducting the study: W.T., B.S., L.A.A.N., H.P.P., S.S.C.R., R.W.M., M.S., J.M.W., I.S., B.M., B.K., W.H., A.L.: collecting and/or interpreting data. W.T., B.S., L.A.A.N., H.P.P., S.S.C.R., R.W.M., M.S., J.M.W., I.S., B.M., B.K., W.H., A.L.: drafting the manuscript. All authors have approved the final draft submitted.

Financial support: Medtronic.

Potential competing interests: W.T., B.S., L.A.A.N., H.P.P., M.S., J.M.W., A.L. reports no conflict of interest. S.S.C.R., R.W.M., B.M., B.K., W.H. has received research grant support from Medtronics. B.K. and I.S. are consultants for Medtronics.

Study Highlights

WHAT IS KNOWN

- ✓ Gastroparesis (GP) has high morbidity and a high economic burden.
- ✓ It is currently difficult to predict which patients may respond to different therapies in suspected GP.

WHAT IS NEW HERE

- ✓ Machine learning model has an acceptable accuracy to predict the response to neuromodulators and/or prokinetics in patients with suspected GP.
- ✓ Gastric transit time based on wireless-motility capsule is comparable with gastric emptying scintigraphy in predicting response to neuromodulators and/or prokinetics.

REFERENCES

1. Jung HK, Choung RS, Locke GR III, et al. The incidence, prevalence, and outcomes of patients with gastroparesis in Olmsted County, Minnesota, from 1996 to 2006. Gastroenterology 2009;136(4):1225–33.
2. Lacy BE, Tack J, Gyawali CP. AGA clinical practice update on management of medically refractory gastroparesis: Expert review. Clin Gastroenterol Hepatol 2022;20(3):491–500.
3. Ingrosso MR, Camilleri M, Tack J, et al. Efficacy and safety of drugs for gastroparesis: Systematic review and network meta-analysis. Gastroenterology 2023;164(4):642–54.
4. Hasler WL, Rao SSC, McCallum RW, et al. Influence of gastric emptying and gut transit testing on clinical management decisions in suspected gastroparesis. Clin Transl Gastroenterol 2019;10(10):e00084.
5. Camilleri M, Kuo B, Nguyen L, et al. ACG clinical guideline: Gastroparesis. Am J Gastroenterol 2022;117(8):1197–220.
6. Schol J, Wauters L, Dickman R, et al. United European Gastroenterology (UEG) and European Society for Neurogastroenterology and Motility (ESNM) consensus on gastroparesis. Neurogastroenterol Motil 2021; 33(8):e14237.
7. Janssen P, Harris MS, Jones M, et al. The relation between symptom improvement and gastric emptying in the treatment of diabetic and idiopathic gastroparesis. Am J Gastroenterol 2013;108(9):1382–91.
8. Vijayvargiya P, Jameie-Oskooei S, Camilleri M, et al. Association between delayed gastric emptying and upper gastrointestinal symptoms: A systematic review and meta-analysis. Gut 2019;68(5):804–13.
9. Vijayvargiya P, Camilleri M, Chedid V, et al. Effects of promotility agents on gastric emptying and symptoms: A systematic review and meta-analysis. Gastroenterology 2019;156(6):1650–60.
10. Pasricha PJ, Yates KP, Nguyen L, et al. Outcomes and factors associated with reduced symptoms in patients with gastroparesis. Gastroenterology 2015;149(7):1762–74.e4.
11. Lee AA, Rao K, Parkman HP, et al. Baseline predictors of longitudinal changes in symptom severity and quality of life in patients with suspected gastroparesis. Clin Gastroenterol Hepatol 2022;20(3):e407–e428.
12. Heo J, Yoon JG, Park H, et al. Machine learning-based model for prediction of outcomes in acute stroke. Stroke 2019;50(5):1263–5.
13. Dong J, Feng T, Thapa-Chhetry B, et al. Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. Crit Care 2021;25(1):288.
14. Ahn JC, Attia ZI, Rattan P, et al. Development of the AI-Cirrhosis-ECG score: An electrocardiogram-based deep learning model in cirrhosis. Am J Gastroenterol 2022;117(3):424–32.
15. Lee AA, Rao S, Nguyen LA, et al. Validation of diagnostic and performance characteristics of the wireless motility capsule in patients with suspected gastroparesis. Clin Gastroenterol Hepatol 2019;17(9): 1770–9.e2.
16. Hasler WL, Lee AA, Moshiree B, et al. Benefits of prokinetics, gastroparesis diet, or neuromodulators alone or in combination for symptoms of gastroparesis. Clin Gastroenterol Hepatol 2024;22(4): 867–77.e12.
17. Revicki DA, Rentz AM, Dubois D, et al. Gastroparesis Cardinal Symptom Index (GCSI): Development and validation of a patient reported assessment of severity of gastroparesis symptoms. Qual Life Res 2004; 13(4):833–44.
18. McCallum RW, Lembo A, Esfandyari T, et al. Phase 2b, randomized, double-blind 12-week studies of TZP-102, a ghrelin receptor agonist for diabetic gastroparesis. Neurogastroenterol Motil 2013;25(11):e705–17.
19. Tougas G, Eaker EY, Abell TL, et al. Assessment of gastric emptying using a low fat meal: Establishment of international control values. Am J Gastroenterol 2000;95(6):1456–62.
20. Kuo B, McCallum RW, Koch KL, et al. Comparison of gastric emptying of a nondigestible capsule to a radio-labelled meal in healthy and gastroparetic subjects. Aliment Pharmacol Ther 2008;27(2):186–96.
21. Kloetzer L, Chey WD, McCallum RW, et al. Motility of the antroduodenum in healthy and gastroparetics characterized by wireless motility capsule. Neurogastroenterol Motil 2010;22(5):527–e117.
22. Waljee AK, Mukherjee A, Singal AG, et al. Comparison of imputation methods for missing laboratory data in medicine. BMJ Open 2013;3(8): e002847.
23. Pawley S. recipeselectors: A collection of steps for feature selection to use with the "recipes" package. 2022.
24. Kuhn MWH. Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles, 2020. Accessed January 10, 2024 https://www.tidymodels.org
25. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1–22.
26. Wright MN, Ziegler A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. J Stat Softw 2017;77(1):1–17.

27. Greenwell BMBB, Boehmke B. Variable importance plots—An introduction to the vip package. R J 2020;12(1):343–66.
28. Staniak M, Biecek P. Explanations of model predictions with live and breakDown packages. R J 2019;10(2):395.
29. Talley NJ, Locke GR, Saito YA, et al. Effect of amitriptyline and escitalopram on functional dyspepsia: A multicenter, randomized controlled study. Gastroenterology 2015;149(2):340–9.e2.
30. Anaparthy R, Pehlivanov N, Grady J, et al. Gastroparesis and gastroparesis-like syndrome: Response to therapy and its predictors. Dig Dis Sci 2009;54(5):1003–10.
31. Yang X, Wen W. Ridge and lasso regression models for cross-version defect prediction. IEEE Trans Reliability 2018;67(3):885–96.
32. Abraham G, Kowalczyk A, Zobel J, et al. Performance and robustness of penalized and unpenalized methods for genetic prediction of complex human disease. Genet Epidemiol 2013;37(2):184–95.
33. Dong Q. Leakage prediction in machine learning models when using data from sports wearable sensors. Comput Intell Neurosci 2022;2022: 5314671.
34. Poldrack RA, Huckins G, Varoquaux G. Establishment of best practices for evidence for prediction: A review. JAMA Psychiatry 2020;77(5): 534–40.
35. Amjad W, Doycheva I, Kamal F, et al. Clinical predictors of symptom improvement failure in gastroparesis. Ann Gastroenterol 2022;35(2):119–26.
36. Bityutskiy LP, Soykan I, McCallum RW. Viral gastroparesis: A subgroup of idiopathic gastroparesis—Clinical characteristics and long-term outcomes. Am J Gastroenterol 1997;92(9):1501–4.
37. Parkman HP, Yates K, Hasler WL, et al. Clinical features of idiopathic gastroparesis vary with sex, body mass, symptom onset, delay in gastric emptying, and gastroparesis severity. Gastroenterology 2011;140(1):101–15.
38. Parkman HP, Van Natta ML, Abell TL, et al. Effect of nortriptyline on symptoms of idiopathic gastroparesis: The NORIG randomized clinical trial. JAMA 2013;310(24):2640–9.
39. Surjanhata B, Brun R, Wilding G, et al. Small bowel fed response as measured by wireless motility capsule: Comparative analysis in healthy, gastroparetic, and constipated subjects. Neurogastroenterol Motil 2018; 30(5):e13268.
40. Barshop K, Staller K, Semler J, et al. Duodenal rather than antral motility contractile parameters correlate with symptom severity in gastroparesis patients. Neurogastroenterol Motil 2015;27(3):339–46.