PREDICT-SM: Development of Machine Learning Models to Support Screening for Undiagnosed Systemic Mastocytosis

Poster Number 530

Daniel S. Herman,[1] Justin Tang,[1] Lindsay Guare,[1] Sayeda Humaira,[1] Ranran Zhang,[1] Daniel Shaheen,[2] Pavle Milutinovic,[3] Elizabeth Hexner,[3] Olajumoke Fadugba[3]

[1]Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; [2]Blueprint Medicines Corporation, Cambridge, MA; [3]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

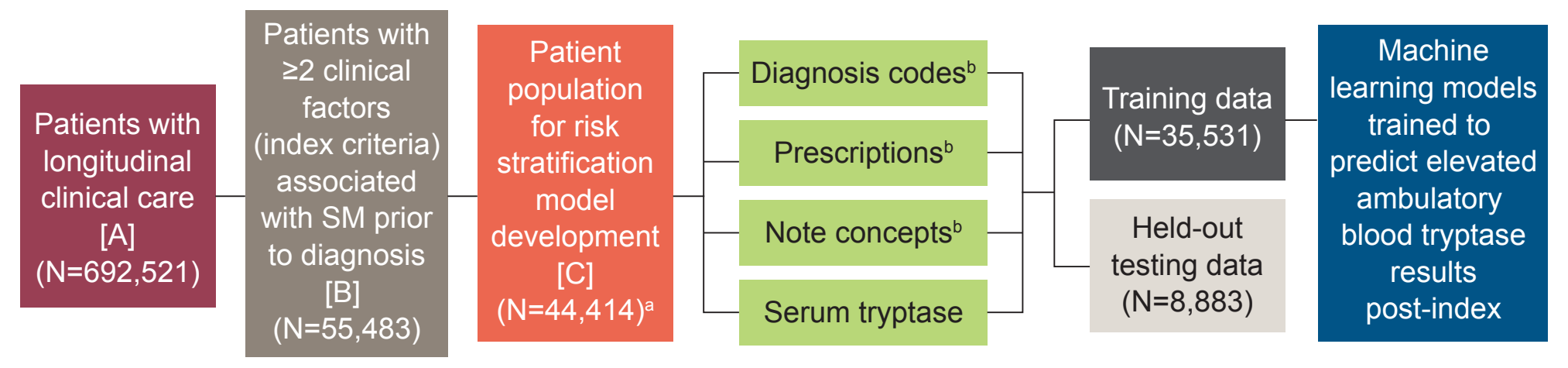
Introduction

- Systemic mastocytosis (SM) is a clonal mast cell disease driven by *KIT* D816V in ~95% of cases,[1–3] characterized by unpredictable symptoms across multiple organ systems that can be debilitating[4–6]
- The major criterion for SM diagnosis is the presence of multifocal mast cell clusters in the bone marrow and/or extracutaneous organs. Minor diagnostic criteria include elevated serum tryptase level, mast cell expression of CD25, CD2 and/or CD30, and presence of activating *KIT* mutations.[4] Clinical manifestations commonly include cutaneous, gastrointestinal, systemic (general weakness/fatigue), neurocognitive symptoms, and life-threatening anaphylaxis[4,6,7] and may have a significant impact on quality of life[8,9]
- The low specificity of symptoms and overall heterogeneity of SM contributes to the diagnostic delays experienced in patients, with delays of up to 9 years from symptom onset to diagnosis observed[10]
- The prevalence of diagnosed mastocytosis has been estimated to be as high as 1 in 5,000 adults[11–14]
- Earlier diagnosis of SM could decrease SM-associated symptoms, improve quality of life, and decrease silent secondary organ damage
- Adoption of electronic health records (EHRs) along with rapid improvement in computational methods has created opportunities to apply machine learning and artificial intelligence (AI) to clinical data to identify patients with underdiagnosed diseases.[15,16] The PREDICT-SM study aims to develop a pragmatic, accurate, and scalable approach to screen for undiagnosed SM by applying AI tools to EHR data
- Here, AI tools were used to train accurate, scalable AI models that could be applied to identify patients who would benefit from SM screening

Methods

- Study cohort [A] was constructed of patients receiving longitudinal clinical care in the Penn Medicine health system with clinical encounters between January 1, 2012, and January 1, 2024 (**Figure 1**)
  - Data from patients who opted out of research within the Penn Medicine health system were not included in this study
- We next filtered for patients with EHR data that included ≥2 clinical factors commonly associated with SM prior to diagnosis (i.e., index criteria) to create a targeted cohort [B]
- The index criteria included 9 diagnosis codes, documented either as a diagnosis for a clinical encounter or listed on a patient's 'problem list', and prescription of medications classified as antihistamines or anaphylaxis therapy agents (**Table 2**)
  - A patient's 'problem list' is a list of overall active medical conditions or issues that should be considered within their individual care plan
- EHR data were extracted from 5 years before each patient met the index criteria, including diagnosis codes (n=261), prescriptions (n=237), and signs or symptoms documented in clinical notes (n=26)
- After the application of exclusion criteria, we used the model development population [C] to develop AI risk stratification models, using logistic regression with Least Absolute Shrinkage and Selection Operator regularization (LR) and histogram-based gradient boosting classification trees (GB). AI models were trained to predict which patients would have a serum tryptase test ordered post-index and a serum tryptase result elevated above the upper limit of the reference interval. Method hyperparameters were tuned by 5-fold cross-validation
- We selected a model interpretive threshold considering the desired use case of identifying patients who should be tested for SM by measuring serum tryptase concentrations and/or blood *KIT* D816V mutations. We targeted a number needed to screen (NNS) of 10, meaning that for every 10 patients the model identified 1 patient that should meet criteria for testing for SM. Estimates are provided assuming that the frequency of patients that should be tested for SM in the model development cohort [C] is 3%

Figure 1. Study Design



Patients with longitudinal clinical care [A] (N=692,521)

Patients with ≥2 clinical factors (index criteria) associated with SM prior to diagnosis [B] (N=55,483)

Patient population for risk stratification model development [C] (N=44,414)[a]

Diagnosis codes[b]

Prescriptions[b]

Note concepts[b]

Serum tryptase

Training data (N=35,531)

Held-out testing data (N=8,883)

Machine learning models trained to predict elevated ambulatory blood tryptase results post-index

[a]Excluding patients with tryptase measured prior to index, less than 6 months of pre-index or post-index data, or age <18 years. [b]Predictors were included from 5 years prior to patients meeting index criteria.
SM, systemic mastocytosis.

Results

- In total, there were 692,521 patients identified with at least 5 visits, including at least 2 visits in primary care, allergy and immunology, dermatology, gastroenterology, or the emergency department (**Table 1**)

Table 1. Description of longitudinal cohort [A] patients grouped by the presence of SM-associated EHR data

Characteristic	Overall [A] (N=692,521)	Index positive: No (N=637,038)	Index positive: Yes [B] 55,483	P-value
Age, median (Q1, Q3)	54.0 (36.0, 78.0)	54.0 (37.0, 69.0)	51.0 (36.0, 65.0)	<0.001
Sex, n (%)				
Female	407,449 (59)	366,597 (58)	40,852 (74)	<0.001
Male	285,023 (41)	270,398 (42)	14,625 (26)	
Unknown	1 (<1)	1 (<1)	0	
Nonbinary	48 (<1)	42 (<1)	6 (<1)	
Race, n (%)				
American Indian or Alaskan Native	1,314 (<1)	1,127 (<1)	187 (<1)	<0.001
Asian	30,875 (4)	28,602 (4)	2,273 (4)	
Black/African American	127,924 (18)	111,814 (18)	16,110 (29)	
East Indian	184 (<1)	173 (<1)	11 (<1)	
Native Hawaiian or other Pacific Islander	841 (<1)	759 (<1)	82 (<1)	
None	15,659 (2)	15,122 (2)	537 (1)	
Patient declined	1,991 (<1)	1,807 (<1)	184 (<1)	
Some other race	23,944 (3)	22,191 (3)	1,753 (3)	
Unknown	22,087 (3)	20,663 (3)	1,424 (3)	
White	467,702 (68)	434,780 (68)	32,922 (59)	
Ethnicity, n (%)				
Hispanic Latino	26,273 (4)	23,933 (4)	2,340 (4)	<0.001
None	4,072 (1)	3,897 (1)	175 (<1)	
Not Hispanic or Latino	658,516 (95)	605,812 (95)	52,704 (95)	
Patient declined	3,554 (1)	3,294 (1)	260 (<1)	
Unknown	106 (<1)	102 (<1)	4 (<1)	
Number of encounters, median (Q1,Q3)	26.0 (12.0, 55.0)	24.0 (12.0, 50.0)	58.0 (29.0, 111.0)	<0.001

EHR, electronic health record; Q, quartile.

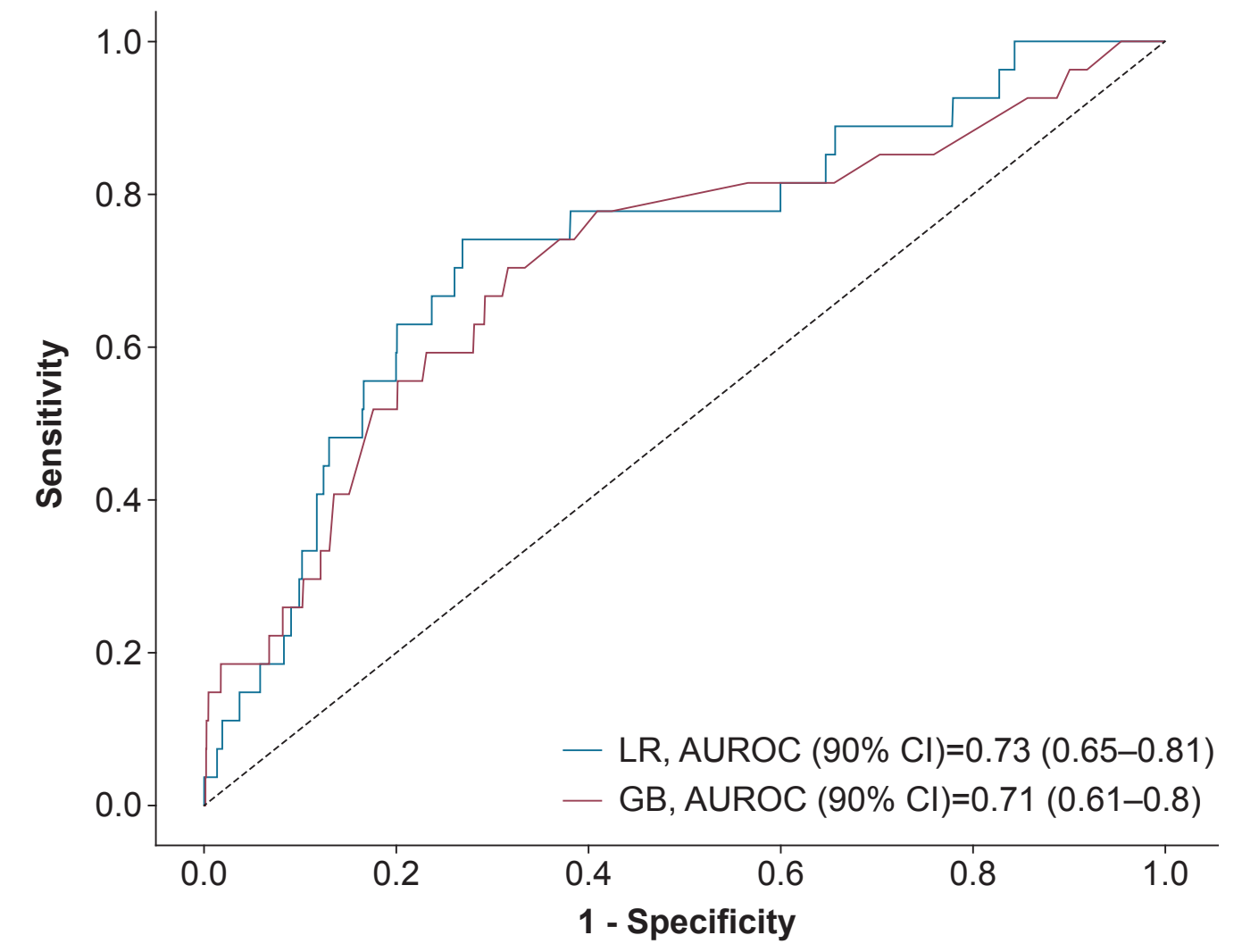
- Within the targeted cohort [B], cetirizine hydrochloride was the most frequent index criterion, followed by loratadine (**Table 2**)
- A total of 44,414 patients were included in the model development cohort [C] because they had some EHR data that could be consistent with SM and did not have tryptase measured prior to meeting the index criteria (**Table 3**)
- In the model development cohort [C], 1,363 patients had serum tryptase ordered and 156 (11%) had elevated serum tryptase results
- In total, there were 572 predictors evaluated using univariate logistic regression. Of these, 30 predictors appeared nominally associated with the compound outcome of tryptase measurement and elevated tryptase results (P<0.025) in univariate analyses (**Table 4**). For the LR model, 6 further predictors were excluded to mitigate feature covariance (Pearson correlation >0.3)
- Within the training data (N=35,531), the LR model performed well at discriminating cases and controls (area under the receiver operating curve [AUROC]=0.82 [90% confidence interval (CI): 0.78–0.85])
- Within the held-out testing data (N=8,883), the LR model demonstrated reasonable discrimination (AUROC=0.73 [90% CI: 0.65–0.81]), which appeared similarly to that of the more complex GB model (**Figure 2**)
- The LR model demonstrated in-testing data sensitivity of 0.48 (90% CI: 0.32–0.64) and an estimated NNS to identify 1 patient that should be tested for SM of 10.9 (90% CI: 6.7–17.6), under the assumption that the frequency of SM testing in this population should be 3% (**Table 5**)
- We used Shapley Additive Explanations (SHAP) to summarize the relative impact of the individual predictors in LR (**Figure 3A**) and GB (**Figure 3B**) model predictions. For most predictors (e.g., flushing), higher values were associated with higher model predicted probabilities for elevated tryptase. However, steroid inhalant prescriptions appeared inversely associated with elevated tryptase, which appeared to be mediated through lower tryptase concentrations rather than less frequent measurement of tryptase (**Table 4**). Loratadine prescriptions also appeared inversely associated with elevated tryptase, but this association appeared to be primarily mediated through less frequent measurement of tryptase
- SHAP values summarize the impact of predictors on AI model outputs by generating an additive feature attribution model. Positive and negative SHAP values indicate a marginal increase and decrease in predictions, respectively. The plots in **Figure 3** depict the distribution of SHAP values relative to the magnitude of each predictor, with each dot representing a single patient

Table 2. Frequency of top index criteria in the targeted cohort [B]

Index criteria, n (%)	Targeted cohort (N=55,483)
Cetirizine HCl	13,755 (25)
Loratadine	8,119 (15)
Fexofenadine HCl	6,492 (12)
Epinephrine	4,966 (9)
Hydroxyzine HCl	4,793 (9)
Diphenhydramine HCl	4,441 (8)
Levocetirizine dihydrochloride	3,836 (7)
L50.9 (urticaria, unspecified)	3,090 (6)
R23.2 (flushing)	2,061 (4)
T78.3XXA (angioedema, initial)	763 (1)
Desloratadine	721 (1)
L50.1 (Idiopathic urticaria)	671 (1)

HCl, hydrochloride.

Figure 2. Model discrimination for held-out testing patients



AUROC, area under the receiver operating curve; CI, confidence interval; LASSO, Least Absolute Shrinkage and Selection Operator; LR, logistic regression with LASSO regularization; GB, gradient-boosting classification tree.

Table 3. Description of the model development cohort [C] grouped by elevated tryptase

Characteristic	Overall [C] (N=44,414)	Index positive: No (N=44,258)	Index positive: Yes (N=156)	P-value
Age, median (Q1, Q3)	53.0 (39.0, 66.0)	53.0 (39.0, 66.0)	59.0 (46.0, 71.0)	<0.001
Sex, n (%)				
Female	33,475 (75)	33,356 (75)	119 (76)	0.956
Male	10,933 (25)	10,896 (25)	37 (24)	
Nonbinary	6 (<1)	6 (<1)	0	
Race, n (%)				
American Indian or Alaskan Native	161 (<1)	160 (<1)	1 (<1)	0.020
Asian	1,742 (4)	1,740 (4)	2 (1)	
Black/African American	13,761 (31)	13,730 (31)	31 (20)	
East Indian	8 (<1)	8 (<1)	0	
Native Hawaiian or other Pacific Islander	64 (<1)	64 (<1)	0	
None	309 (1)	309 (1)	0	
Patient declined	136 (<1)	136 (<1)	0	
Some other race	1,343 (3)	1,340 (3)	3 (2)	
Unknown	937 (2)	935 (2)	2 (1)	
White	25,953 (58)	25,836 (58)	117 (75)	
Ethnicity, n (%)				
Hispanic Latino	1,842 (4)	1,840 (4)	2 (1)	0.433
None	113 (<1)	113 (<1)	0	
Not Hispanic or Latino	42,266 (95)	42,113 (95)	153 (98)	
Patient declined	191 (<1)	190 (<1)	1 (1)	
Unknown	2 (<1)	2 (<1)	0	
Tryptase, median (Q1, Q3)	4.7 (3.4, 6.3)	4.4 (3.2, 5.6)	11.0 (9.2, 15.5)	<0.001
Allergy visits, n (%)	3,858 (9)	3,818 (9)	40 (26)	<0.001
Dermatology visits, n (%)	11,187 (25)	11,136 (25)	51 (33)	0.038
Family Practice visits, n (%)	12,049 (27)	12,022 (27)	27 (17)	0.008
Gastroenterology visits, n (%)	7,305 (16)	7,264 (16)	41 (26)	0.001
Gerontology visits, n (%)	338 (1)	338 (1)	0	0.636
Hematology/oncology visits, n (%)	4,039 (9)	4,012 (9)	27 (17)	0.001
Internal medicine visits, n (%)	22,088 (50)	22,017 (50)	71 (46)	0.329
Pediatrics visits, n (%)	291 (1)	291 (1)	0 (0)	0.630

Table 4. Univariate logistic regression in model development cohort [C]

	Tryptase ordered (I)		Tryptase elevated (II)		Tryptase ordered and elevated (III)			
Predictor	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Controls, %[a]**	**Cases, %**	**Coefficient**	**P-value**
Flushing	4.943	2.79E-30	2.599	1.33E-02	9.41	25.58	5.702	6.45E-14
Urticaria pigmentosa	22.564	3.71E-07	9.849	2.55E-03	0.04	6.2	20.296	6.94E-09
Anaphylaxis	2.928	9.48E-44	0.52	4.57E-01	15.24	41.09	2.973	1.25E-08
D47.01 (cutaneous mastocytosis)	2.704	1.54E-04	2.809	1.51E-02	0.01	2.33	3.6	3.90E-06
L50.9 (urticaria, unspecified)	0.239	3.66E-26	0.018	8.14E-01	11.71	30.23	0.161	4.42E-06
Hypotension	2.893	2.28E-08	2.35	8.01E-02	7.19	14.73	4.007	1.91E-05
T78.1XXA (adverse food reaction)	0.654	9.15E-32	–0.027	8.82E-01	2.4	10.85	0.533	2.49E-05
Itching	0.519	2.88E-02	2.338	4.99E-04	73.42	84.5	2.051	8.87E-05
Anaphylaxis therapy agents	0.304	5.87E-28	–0.004	9.67E-01	14.77	30.23	0.277	1.54E-04
Pressors	0.304	2.94E-28	–0.009	9.29E-01	14.84	30.23	0.275	1.75E-04
Epinephrine	0.304	3.00E-28	–0.009	9.29E-01	14.84	30.23	0.275	1.75E-04
Loratadine	–0.433	4.10E-14	–0.17	2.97E-01	34.43	15.5	–0.716	3.09E-04
Zafirlukast	0.39	6.27E-07	0.186	2.49E-01	0.27	3.1	0.388	5.53E-04
T78.3XXD (angioedema, subsequent)	1.189	1.15E-17	0.049	8.78E-01	0.86	3.1	1.054	6.03E-04
T88.6XXA (anaphylactic reaction due to adverse effect of correct drug)	2.768	9.19E-05	1.329	2.79E-01	0.02	0.78	3.572	6.90E-04
Allergy status to other antibiotic agents	1.12	2.00E-05	0.524	2.78E-01	0.17	1.55	1.464	9.54E-04
Pruritis	1.328	3.99E-02	3.617	2.33E-02	10.11	16.28	3.314	1.65E-03
Syncope	0.967	1.24E-03	1.567	5.89E-02	32.28	44.96	2.006	2.96E-03
Age	–0.01	1.07E-07	0.029	3.44E-07	53	57	0.015	3.00E-03
T78.3XXA (angioedema, initial)	0.252	1.66E-18	–0.032	7.24E-01	3.54	11.63	0.12	4.90E-03
Ibandronate sodium	0.02	8.79E-01	1.333	2.76E-01	0.41	1.55	0.326	5.66E-03
Z87.2 (diseases of skin)	0.15	3.12E-02	0.991	1.01E-01	0.62	2.33	0.335	6.77E-03
D72.19 (eosinophilia)	0.289	4.45E-03	0.316	1.78E-01	0.14	0.78	0.352	7.96E-03
Epinephrine HCl	0.546	4.55E-02	1.144	1.22E-01	0.03	0.78	0.746	9.45E-03
Olopatadine HCl	0.153	2.43E-01	–0.691	4.21E-01	1.06	1.55	0.463	1.06E-02
L50.1 (idiopathic urticaria)	0.257	5.57E-08	0.022	9.14E-01	1.68	9.3	0.16	1.39E-02
Steroid inhalants	–0.008	7.41E-01	–0.981	6.74E-03	12.56	3.88	–0.784	1.50E-02
Cimetidine	0.185	2.60E-01	1.147	3.44E-02	0.3	1.55	0.448	1.79E-02
Miscellaneous endocrine	0.003	9.44E-01	0.161	4.59E-02	3.34	6.2	0.137	2.04E-02
Bone density regulators	0.003	9.44E-01	0.161	4.59E-02	3.34	6.2	0.137	2.04E-02

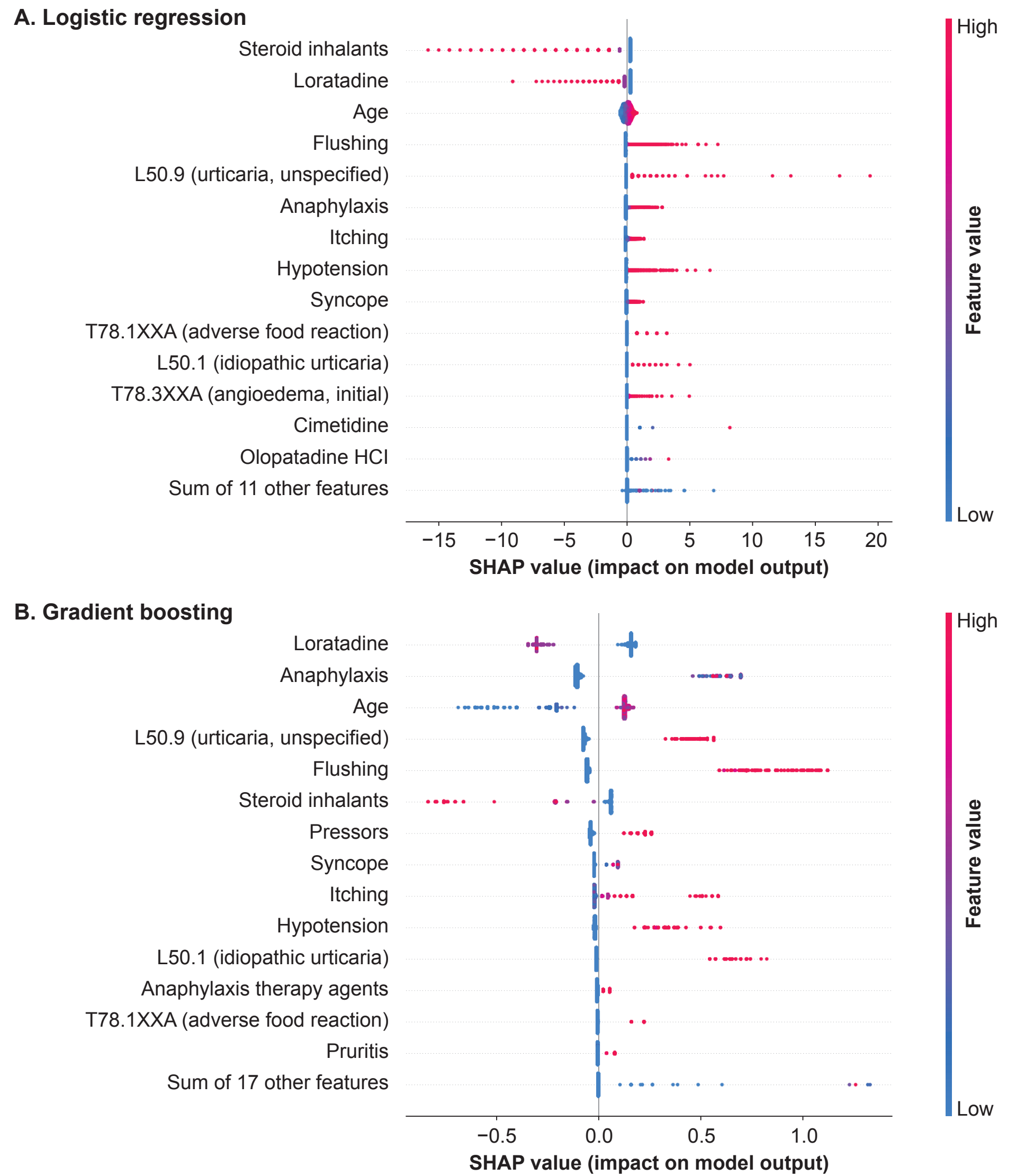
Coefficients (and associated P-values) from univariate logistic regression to predict who in cohort (C) had a tryptase order placed (I), in the subset of patients with an order placed who had an elevated tryptase result (II), and who in the full cohort (C) had a tryptase order placed and the result was elevated (III).
[a]The percent of patients who had at least observation of the predictor.
Note that coefficient magnitudes cannot be directly compared across predictor types because of differences in predictor scaling.

Table 5. Model classification performance of the LR model in held-out testing

	Estimate	SE	90% CI
Sensitivity	0.48	0.10	0.32–0.64
Precision	0.10	0.03	0.06–0.15
NNS	10.9	3.7	6.7–17.6

CI, confidence interval; NNS, number needed to screen; SE, standard error.

Figure 3. Explanations of the impact of predictors in AI models using SHAP



SHAP, Shapley Additive Explanations.

Conclusions

- The developed interpretable AI model appears to identify patients who should be screened for SM
- Diagnosis codes (e.g., D47.01), medication prescriptions (e.g., epinephrine), and concepts in clinical notes (e.g., flushing) contribute complementary information for the AI models
- This approach, with further refinements, could ultimately be applied clinically to identify patients who are currently undiagnosed
- Future work is needed to:
  - Improve the extraction of clinical concepts from notes
  - Bridge the gap between predicting tryptase elevation and identifying patients that should be screened for SM
  - Improve the AI models' specificity and generalizability

References

1. Kristensen T et al. *Am J Hematol*. 2014;89:493–498; 2. Ungerstedt J et al. *Cancers*. 2022;14:3942; 3. Garcia-Montero AC et al. *Blood*. 2006;108:2366–2372; 4. Pardanani A. *Am J Hematol*. 2023;98:1097–1116; 5. Jennings S et al. *J Allergy Clin Immunol Pract*. 2014;2:70–76; 6. Valent P et al. *Int J Mol Sci*. 2019;20:2976; 7. Hartmann K et al. *J Allergy Clin Immunol*. 2016;137:35–45; 8. van Anrooij B et al. *Allergy*. 2016;71:1585–1593; 9. Mesa RA et al. *Cancer*. 2022;128: 3691–3699; 10. Jennings SV et al. *Immunol Allergy Clin North Am*. 2018;38:505–525; 11. Bergstrom A et al. *Acta Oncol*. 2024;63:44–50; 12. Brockow K. *Immunol Allergy Clin North Am*. 2014;34:283–295; 13. van Doormaal JJ et al. *J Allergy Clin Immunol*. 2013;131:1429–1431.e1421; 14. Cohen SS et al. *Br J Haematol*. 2014;166:521–528; 15. Cohen AM et al. *PLoS One*. 2020;15:e0235574; 16. Zhang L et al. *J Am Med Inform Assoc*. 2020;27:119–126.

Acknowledgments

Medical writing support was provided by Rachel O'Meara, PhD, and Travis Taylor, BA, of Paragon. Funded by Blueprint Medicines Corporation. The sponsor reviewed and provided feedback on the poster. However, the authors had full editorial control and provided final approval of all content.

Conflicts of interest/disclosures

Daniel S Herman has participated in investigator-initiated research sponsored by Blueprint Medicines Corporation. Full disclosures for all authors are available upon request at medinfo@blueprintmedicines.com.

Poster available for download at:

Presented at the AAAAI/WAO Joint Congress, February 28–March 3, 2025 San Diego, CA. Please contact medinfo@blueprintmedicines .com for permission to reprint or distribute.