Project BEACON: Developing a Predictive Model To Accelerate Diagnosis of Indolent Systemic Mastocytosis

Poster Number 517

Percy Van Crocker, Jr,[1] Linde Finsrud Wilson,[2] Amber King,[3] Fatima Scipione,[3] Ted Brasfield,[4] Princess U. Ogbogu[5]

[1]Agonist Health, Chicago, IL, USA; [2]HealthQuest Advisors, Chicago, IL, USA; [3]Blueprint Medicines Corporation, Cambridge, MA, USA; [4]EveryLife Foundation for Rare Diseases, Washington DC, USA; [5]University Hospitals Rainbow Babies & Children's Hospital & Case Western Reserve University, Cleveland, OH, USA

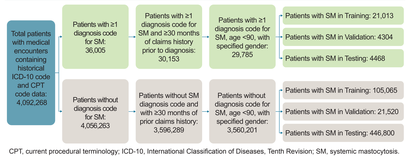
Introduction

- Systemic mastocytosis (SM) is a clonal mast disease with a reported prevalence of up to 1 in 5,000; however, underdiagnosis likely results in underestimation of the true disease prevalence[1]
- SM is characterized by non-specific symptoms that are often evaluated in isolation, resulting in misclassification and prolonged diagnostic delays, particularly for indolent systemic mastocytosis (ISM), where median time from symptom onset to diagnosis is up to 9 years[2]
- During delays in diagnosis, the symptoms of ISM impose a significant burden on daily life and healthcare utilization (including emergency care),[3–5] often for patients with ISM impacted by social determinants of health and limited access to specialist care[6–8]
- Existing models to support earlier diagnosis of SM rely on practice-specific electronic health record data which may limit generalizability and overlook early clinical signals prior to entry into these systems[9,10]
- We focused our predictive model on identifying the risk of SM in such signals as experienced in early-stage populations, where ISM is expected to be prevalent
- We used a supervised learning approach to train a model on approximately 4 million patients where diagnosis status was already known. The model identified combinations of symptoms and procedures in patient medical histories that most strongly predicted eventual SM diagnosis, then translated these learned patterns into a highly interpretable, rules-based algorithm that weighted each clinical feature to calculate the probability of disease
- The predictive model was developed for use in settings where patients often present early in the disease course and where SM is frequently underrecognized, such as primary care, dermatology, community oncology, and allergy and immunology practice settings
- This work focused on helping clinicians "connect the dots" to identify and establish early diagnoses of SM in patients who repeatedly present with multiple unexplained symptoms

Methods

- Our model was constructed using medical claims data from Komodo Health™ representing approximately 4 million patients and including information on patient demographics, diagnoses, and procedures
- The model features were defined by SM clinical experts and included diagnosis and procedure codes that represent early and/or well-characterized clinical identifiers of SM
- Patients were filtered based on the presence of ≥30 months of enrollment history to ensure completeness of claims history
- Patients were split into SM and non-SM cohorts based on the presence of at least one medical claim with an ICD-10 code that corresponds to or is highly associated with SM (D47.02 [Systemic mastocytosis], C94.30 [Mast cell leukemia, not having achieved remission], C94.31 [Mast cell leukemia, in remission], C94.32 [Mast cell leukemia, in relapse], and C96.21 [Aggressive systemic mastocytosis])
- Details of SM-associated symptoms and procedures were extracted from claims made between 6 and 30 months prior to the index date
  - For target patients, the index date was defined as the date of the first SM diagnosis
  - For non-target patients, the index date was defined as the most recent claim
- A 6-month lag period was applied to patient claims used for model training to exclude claims that may have originated after SM screening or diagnosis but prior to the patient's index date; this design filtered the available data to focus on patient encounters unrelated to screening or diagnosis since they were the most suitable to predict SM risk in an undiagnosed population
- The target and non-target patient cohorts were randomly split into three sets: training (68% of patients), validation (17% of patients), and testing (15% of patients) (**Figure 1**)
  - Model training was the process of optimizing the coefficients that describe the relationship between the symptoms, procedures, and SM
  - Model validation was the process of evaluating the model in parallel with training to increase the model's generalizability to unseen data
  - Model testing was used to assess the model's performance with metrics in an unbiased way by using the model equation to make predictions on unseen data
- For the training and validation sets, the model assumes a disease prevalence of 1:5 (ratio of patients with an SM code to those without an SM code), reducing extreme class imbalance to allow the model algorithm to learn patterns from the data. For the testing set, we set a ratio of 1:100 to more closely reflect the real-world prevalence of SM. Stratified random sampling based on age and gender was used to match the cohort with an SM code to the cohort without an SM code

Figure 1. Patient cohort selected for model training, validation, and testing



CPT, current procedural terminology; ICD-10, International Classification of Diseases, Tenth Revision; SM, systemic mastocytosis.

Results

- We identified 29,785 patients with at least one claim containing an SM diagnosis code and 3,560,201 patients without an SM claim who met the selection criteria; baseline age and gender distributions at the index date were evaluated prior to splitting the data for training, validation, and testing (**Table 1**)

Table 1. Baseline characteristics

Age group	Gender	Patients with SM	Patients without SM
0–15	F	818	65,827
0–15	M	977	76,188
16–30	F	2706	138,009
16–30	M	768	70,642
31–45	F	5108	308,040
31–45	M	1392	135,041
46–60	F	5608	440,567
46–60	M	2188	267,683
61–75	F	4567	576,998
61–75	M	2720	484,474
76–90	F	1616	556,996
76–90	M	1317	439,736
Total		**29,785**	**3,560,201**

F, female; M, male.

- Applying the model to the test set resulted in a precision of 0.157, a sensitivity of 0.2001, and a number needed to screen of 6.353 (**Table 2**); the receiver operating characteristic (ROC) curve and precision-recall curve for the test set indicate the model demonstrated reasonable discrimination between patients with SM codes and patients without SM codes (**Figure 2**, **Figure 3**)

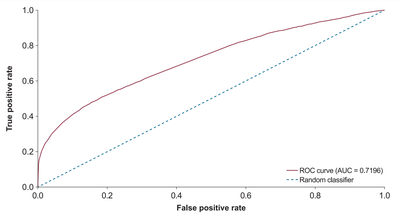
Table 2. Metrics and values following application of the model to the test set

Metric	Value	Description
Accuracy	0.982	Proportion of all predictions (patients with and without an SM code) that the model classified correctly
Precision	0.157	Of all predicted patients with an SM code, the proportion that were observed to have an SM code (measures false-positive control)
Recall/sensitivity	0.2001	Of all patients with an SM code, the proportion the model correctly identified
Number needed to screen	6.353	For every ~six patients that meet the threshold of a high risk of SM within the model, one patient is expected to be diagnosed with SM by a physician

Category	Value	Description
True negatives	442,014	Patients without an SM code correctly predicted by the model to be patients without an SM code
False positives	4786	Patients without an SM code incorrectly predicted by the model to be patients with an SM code
False negatives	3574	Patients with an SM code incorrectly predicted by the model to be patients without an SM code
True positives	894	Patients with an SM code correctly predicted by the model to be patients with an SM code

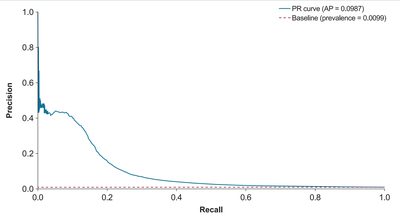
- We used Shapley Additive Explanations (SHAP) to quantify the relative strength of predictors. Mean absolute SHAP values (**Figure 4**) summarize the average magnitude of each predictor's contribution to the model predictions across patients. Diagnoses of unspecified anemia (D64.9) and unspecified headache (R51.9) showed the largest mean absolute SHAP values, indicating the greatest overall influence on predictions in the cohort

Figure 2. Model discrimination between patients with and without SM (ROC curve)



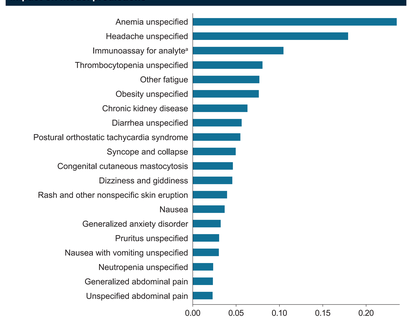
AUC, area under the curve; ROC, receiver operating characteristic.

Figure 3. Precision *vs* recall: Illustrates inverse relationship between precision and recall in the test cohort



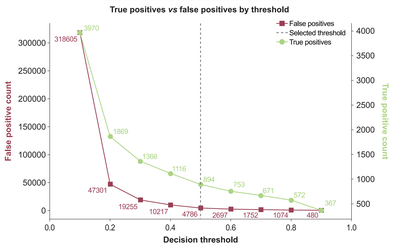
AP, average precision.

Figure 4. Mean absolute SHAP values: Key drivers with the strongest overall impact on model predictions



*CPT 83520. This code includes serum tryptase among other analytes.
SHAP, Shapley Additive Explanation.

Figure 5. True positives and false positives as decision threshold changes



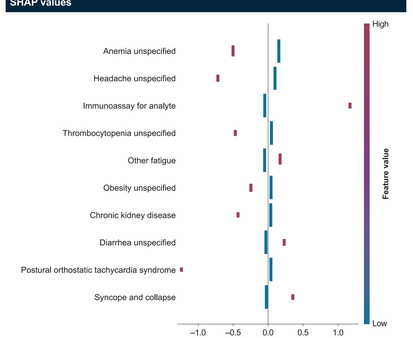
- As the decision threshold increases, the quantity of true positives and false positives both decrease with false positives decreasing at a faster rate (**Figure 5**)
- Our model used a decision threshold of 0.5 for making predictions, representing the minimum predicted probability required for a positive SM classification, to balance the identification of patients with SM codes against the preservation of robust overall precision

Table 3. Feature coefficients: Key drivers of increased/decreased risk for individual patients

Rank	Feature	Coefficient	Absolute value of coefficient	Odds ratio	Effect
1	Congenital cutaneous mastocytosis	3.495	3.495	32.949	Increases risk
2	Anaphylaxis	2.022	2.022	7.557	Increases risk
3	Cutaneous mastocytosis	1.589	1.589	4.899	Increases risk
4	Eosinophilia	1.491	1.491	4.444	Increases risk
5	Postural orthostatic tachycardia syndrome	–1.278	1.278	0.279	Decreases risk
6	Immunoassay for analyte (CPT 83520)	1.217	1.217	3.376	Increases risk
7	Migraine unspecified	–1.211	1.211	0.298	Decreases risk
8	Post-viral and related fatigue syndromes	1.112	1.112	3.041	Increases risk
9	Osteoporosis with current pathological fracture	0.987	0.987	2.683	Increases risk
10	Hypermobility syndrome	0.849	0.849	2.338	Increases risk
11	Headache unspecified	–0.816	0.816	0.442	Decreases risk
12	Dyspnea	–0.764	0.764	0.466	Decreases risk
13	Anemia unspecified	–0.652	0.652	0.521	Decreases risk
14	Flushing	0.634	0.634	1.886	Increases risk
15	Colon screening	–0.604	0.604	0.546	Decreases risk
16	Neutropenia	–0.581	0.581	0.559	Decreases risk
17	Vascular headache not elsewhere classified	0.538	0.538	1.712	Increases risk
18	Thrombocytopenia	–0.518	0.518	0.596	Decreases risk
19	Disorder of bone unspecified	0.504	0.504	1.655	Increases risk
20	Chronic kidney disease	–0.469	0.469	0.626	Decreases risk

- Among the 20 most predictive features, 11 were associated with an increased likelihood of SM, whereas nine were associated with a decreased likelihood (**Table 3**). The model employed least absolute shrinkage and selection operator (LASSO) to prevent overfitting and reduce model complexity. LASSO regression reduces the coefficients of less informative predictors to 0 in order to minimize noise during training and improve predictions

Figure 6. Distribution of symptom and procedure effects on SM risk based on SHAP values



- We evaluated the contribution of individual predictors using SHAP values summarized in a global SHAP summary plot (**Figure 6**). Features are ordered by decreasing mean absolute SHAP value, reflecting their average influence on model predictions across individuals, whether positive or negative. Positive SHAP values indicate an increased predicted likelihood of SM classification, whereas negative values indicate a decreased likelihood. Among the predictors with the largest mean absolute SHAP values, claims including codes for unspecified headache and unspecified anemia were associated with a significant decrease in the risk of SM relative to baseline, while the presence of CPT 83520 (immunoassay for analyte) increased risk
- The waterfall SHAP plot identifies the primary features contributing to the predicted risk of SM for an illustrative individual patient from our claims dataset (**Figure 7**). Because class-imbalance adjustments were applied during model training and artificially increased the assumed prevalence of SM, the model's predicted probabilities do not reflect real-world disease risk. However, SHAP values remain informative for interpreting model predictions at the patient level, as they quantify how the presence or absence of specific features contributes to patient classification
- A trade-off exists in machine learning (ML) classification models where higher decision thresholds result in improved precision and weaker sensitivity (**Figure 8**)
- As the decision threshold is lowered, the model exhibits an expected sharp decline in precision. For example, at a threshold of 0.3, precision decreases to 0.07 while sensitivity increases to 0.31, reflecting a high rate of false-positive predictions in a cohort with high class imbalance

Figure 7. Most influential symptoms and procedures via SHAP values: Example in an individual patient

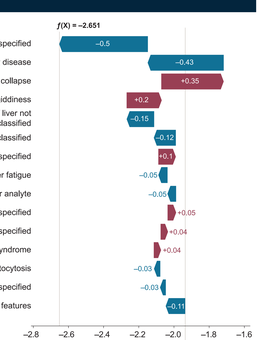
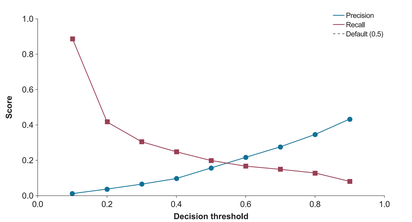


Figure 8. Impact of decision threshold on precision and recall



Conclusions

- Earlier identification of patients with SM through ML could substantially shorten the time to diagnosis, reducing years of unexplained symptoms, unnecessary interventions, and associated impact on quality of life
- By ensuring earlier, comprehensive testing and/or streamlined referral to SM medical experts, this model can interrupt years of missed or incorrect diagnosis, prevent avoidable disease progression, and reduce the long-term clinical, emotional, and economic burden on patients and the healthcare system
- The ML model demonstrates the predictive value of longitudinal care patterns captured in claims data for identifying patients with SM earlier
- Deployment of this predictive model could shorten diagnostic delays, decrease avoidable emergency room visits and hospitalizations, and lower the total cost of care
- Future work is needed to determine:
  - The model's generalizability at the level of a health system with a substantial patient population and sufficient medical data
  - The model's value as a decision support tool for clinicians in a real-world practice setting with the goal of reducing prolonged diagnosis uncertainty and providing patients with solutions earlier

References

1. Bergstein A et al. *Acta Oncol*. 2024;63:44–50; 2. Jennings S et al. *Immunol Allergy Clin North Am*. 2018;38:505–525; 3. Tse KY et al. *J Allergy Clin Immunol Glob*. 2024;27:100316; 4. Mesa RA et al. *Cancer*. 2022;128:3691–3699; 5. Mukherjee S et al. Presented at ASH 2023. Presentation 75; 6. Vrtkaga K et al. *J Brown Hosp Med*. 2025;4:138072; 7. Lim A et al. *BMC Public Health*. 2024;24:1153; 8. Jennings SV et al. *J Allergy Clin Immunol Pract*. 2022;10:2039–2051; 9. Herman DS et al. Presented at AAAAI/WAO 2025 Joint Congress. Poster 530; 10. Dranitsaris G et al. Presented at SOHO 2023. Poster MPN-073

Conflicts of interest/disclosures

Percy Van Crocker has no disclosures to share. For full author disclosures, please contact medinfo@blueprintmedicines.com.

Scan the QR code to view our Interactive Poster

Acknowledgments

We would like to acknowledge Ethan Royce and Alex Armour of HealthQuest Advisors for their assistance reviewing and editing this poster. Medical writing support was provided by Linda Brown, BSc (Hons), and editorial support was provided by Travis Taylor, BA, of the Prime Group of Companies (Knutsford, UK), and was funded by Blueprint Medicines Corporation, a wholly-owned subsidiary of Sanofi. Responsibility for all opinions, conclusions, and data interpretation lies with the authors.