Utility of Large Language Models to Quantify Diagnostic Delays in Systemic Mastocytosis: A Real-World Study

Poster Number 544

Cecilia Arana Yi,[1] Hunter Mills,[2] Syed Arsalan Ahmed Naqvi,[1] Muhammad Ali Khan,[1] Umair Ayub,[1] Ahmed Ibrahim,[1] Shadera Slatter,[2] Lisa Sproat,[1] Nandita Khera,[1] Kate Freeman,[3] John Lewis,[3] Animesh Pardanani,[4] Ayalew Tefferi,[4] Michelle Elliot,[4] Aref Al-Kali,[4] Thanai Pongdee,[5] Joseph Butterfield,[5] Candido Rivera,[6] Ewa Wysokinska,[6] Daniel Shaheen,[7] Irbaz Bin Riaz,[1] Vivek A. Rudrapatna[8]

[1]Division of Hematology, Mayo Clinic, Phoenix, AZ; [2]Bakar C USA Computational Health Sciences Institute, University of California, San Francisco, CA; [3]Division of Allergic Diseases, Mayo Clinic, Phoenix, AZ; [4]Division of Hematology, Mayo Clinic, Rochester, MN; [5]Division of Allergic Diseases, Mayo Clinic, Rochester, MN; [6]Division of Hematology, Mayo Clinic, Jacksonville, FL; [7]Blueprint Medicines Corporation, Cambridge, MA; [8]Division of Gastroenterology, Department of Medicine University of California, San Francisco, CA, USA

Introduction

- Systemic mastocytosis (SM) is a rare clonal mast cell disorder driven by the *KIT* D816V mutation in ~95% of patients[1–3]
- Historically, the prevalence of SM has been estimated at 1 in 10,000 people,[4–6] although a recent study suggests that it could affect up to 1 in 5,000 people[7]
- SM is characterized by chronic, non-specific, and often debilitating symptoms that can significantly impact quality of life[8–11]
- The heterogenous and non-specific nature of symptoms frequently results in significant diagnostic delays, with an average delay of approximately 6 years from symptom onset[8]
- Large language models (LLMs) can assist in reducing diagnostic delays by detecting SM-related symptoms in electronic health records, enabling earlier identification and intervention[12]

Methods

- Patients with SM were identified by the International Classification of Diseases (ICD) codes at the Mayo Clinic from 2017 onwards
- The following ICD codes were used to identify the cases
  - ICD-10: D47.02, C96.21, C94.30, C96.22
  - ICD-9: 202.60, 238.5
- The identified patients had their SM manually confirmed by a hematologist (SM expert) using chart review
- The study was institutional review board–approved at the University of California San Francisco (#22-37116) and at the Mayo Clinic (#22-007433)
- Eligible patients with a manually confirmed diagnosis of SM were randomly sampled into development and test sets for iterative prompt engineering and evaluation, respectively
- The state-of-the-art LLM, generative pretrained transformer (GPT)-4, was used to extract symptoms related to SM from clinical notes prior to the date of diagnosis
- A zero-shot structured prompt was used to call the GPT-4 LLM
- The LLM-based symptom assessment was benchmarked against the manually curated variables for accuracy
- Median time from each preceding symptom to disease diagnosis was computed separately for LLM-based as well as the manual approach using the Kaplan-Meier method

Results

Baseline characteristics

- A total of 301 patients with 560 eligible clinical notes were included in this study
- The development set included 5 patients with 53 clinical notes, and the test set included 296 patients with 507 clinical notes
- The LLM-based approach identified SM-related symptoms in 301 patients (**Table 1**)
- Diagnosis and symptom details, and prior to SM diagnosis medication use are presented in **Table 2** and **Table 3**, respectively
- *KIT* D816V was detected in 62 of 78 (79%) patients tested
  - *KIT* D816V mutations were detected by allele-specific droplet digital polymerase chain reaction

Table 1. Baseline characteristics

Characteristic	Overall (N=301)
Gender, n (%)	
Female	172 (57)
Male	129 (43)
Race, n (%)	
White	277 (92)
Asian	0
Native American/Alaskan/Hawaiian	3 (1)
Black/African American	5 (2)
Other	5 (2)
Unknown	11 (4)
Ethnicity, n (%)	
Hispanic/Latino	13 (4)
Area Deprivation Index	
National, median (IQR)	48.0 (28.0–68.0)
State, median (IQR)	5.0 (3.0–8.0)
Median driving distance to Mayo Clinic, miles (IQR)	161.9 (72.1–352.7)
Patients with insurance, n (%)	
Commercial	140 (47)
Medicare	39 (13)
Medicaid	0
Self-pay or other	0
Unknown	122 (41)
Age at the time of diagnosis, median (IQR)	57.3 (46.7–68.0)
Year of SM diagnosis, median (IQR)	2021 (2019–2023)
Patients with SM diagnosis, n (%)	
Before 2020	115 (38)
After 2020 (COVID-19 era)	186 (62)

IQR, interquartile range; SM, systemic mastocytosis.

Performance for symptom extraction

- The LLM-based approach demonstrated high accuracy in detecting SM-related symptoms from clinical notes, with accuracy of 100% and 99% for the development and test external sets, respectively
- The LLM-based approach demonstrated a median time of 1.8 years from the onset of symptoms to definitive diagnosis
- The manual approach demonstrated a median time of 1.7 years from the onset of symptoms to definitive diagnosis (**Figure 1**)

Table 2. Diagnosis and symptom details

Variable	Overall (N=301)
Type of SM diagnosis	
Indolent SM	230 (76)
Smoldering SM	7 (2)
Aggressive SM	18 (6)
SM with associated hematologic neoplasm	40 (13)
Mast cell leukemia	4 (1)
Mast cell sarcoma	2 (1)
Not otherwise specified	0
Diagnostic journey	
Dermatological symptoms prior to diagnosis	188 (63)
Preceding cutaneous mastocytosis	39 (13)
Tested for *KIT* D816V	78 (26)
Positive for *KIT* D816V	62 (21)
Elevated serum tryptase (>11.5 ng/mL)	14 (5)
Symptom history	
Anaphylaxis	38 (13)
Fatigue	91 (30)
Lightheadedness, syncope/fainting	66 (22)
Skin symptoms[a]	124 (41)
Gastrointestinal symptoms[b]	150 (50)
Neuropsychiatric symptoms[c]	110 (37)
Cardiovascular symptoms[d]	33 (11)
Pulmonary symptoms[e]	71 (24)
Musculoskeletal symptoms[f]	113 (38)
Nasal/throat symptoms[g]	78 (26)

[a]Flushing, itching, and hives. [b]Pain, diarrhea, nausea, vomiting, bloating, and reflux. [c]Headache, brain fog, cognitive dysfunction, anxiety, depression, and inability to concentrate. [d]Palpitations, chest pain, and blood pressure instability. [e]Cough, wheezing, and shortness of breath. [f]Bone/muscle pain, osteopenia, osteoporosis, and osteosclerosis. [g]Congestion, throat itching, and swelling

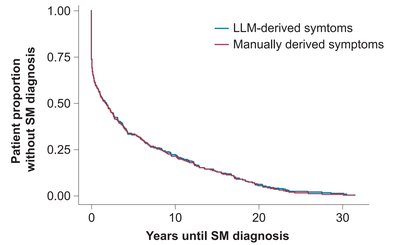
Predictors of diagnostic delays

- Univariate regression analyses were performed to assess predictors of diagnostic delays
- Patients >50 years of age, White race, area deprivation index 28–68 (interquartile range) *vs* <28 (lower quartile), and shorter driving distance were more likely to experience diagnostic delays (**Table 4**)
- Patients without anaphylaxis, fatigue, or dermatologic, nasal/throat (upper respiratory), pulmonary (lower respiratory), or abdominal symptoms were more likely to experience diagnostic delays than those with these symptoms

Table 3. Medication use prior to diagnosis of SM

Medication	Overall (N=301)
Cromolyn	7 (2)
Montelukast	8 (3)
Omalizumab	3 (1)
Interferon	1 (<1)
Tyrosine kinase inhibitors	
Avapritinib	1 (<1)
Imatinib	1 (<1)
H1 antihistamines	
Diphenhydramine	17 (6)
Doxepin	7 (2)
Fexofenadine	31 (10)
Ketotifen	1 (<1)
Loratadine	36 (12)
H2 antihistamines	
Cimetidine	4 (1)
Famotidine	20 (7)
Corticosteroids	
Budesonide	14 (5)
Dexamethasone	15 (5)
Hydrocortisone	11 (4)
Prednisone	8 (3)
Methylprednisolone	5 (2)
Fludrocortisone	3 (1)

Figure 1. Kaplan-Meier curves showing time from initial symptom onset to definitive SM diagnosis with symptom identification



LLM, large language model; SM, systemic mastocytosis.

Table 4. Univariate regression analyses highlighting significant predictors of diagnostic delays

Variable	Delta (years)	P-value	n
Demographic details			
Age: >50 years (*vs* age ≤50 years)	2.4	0.004	189
Gender: female (*vs* male)	0.32	0.70	172
Race: White (*vs* non-White)	4.1	0.03	277
Ethnicity: Hispanic/Latino (*vs* non-Hispanic/Latino)	–4.9	0.12	13
ADI: IQR (*vs* LQ)	7.2	1.8×10^{-4}	105
ADI: UQ (*vs* LQ)	–0.9	0.53	50
Driving distance from Mayo, miles: IQR (*vs* LQ)	–3.7	1.4×10^{-4}	146
Driving distance from Mayo, miles: UQ (*vs* LQ)	-6.3	3.4×10^{-8}	73
Symptom history[a]			
Nasal/throat symptoms	–1.74	0.04	78
Musculoskeletal symptoms	–1.37	0.09	113
Pulmonary symptoms	–1.87	0.04	71
Cardiovascular symptoms	–1.02	0.41	33
Dermatologic symptoms	–2.62	0.002	124
Anaphylaxis	–3.87	7×10^{-4}	38
Fatigue	–3.11	2×10^{-4}	91
Lightheadedness	–0.62	0.51	66
Abdominal symptoms	–2.42	0.003	118
Neuropsychiatric symptoms	–0.97	0.24	110

[a]All symptom history comparisons are *versus* the group of patients without a history of the corresponding symptom. ADI, area deprivation index; LQ, lower quartile; UQ, upper quartile.

Table 5. Multivariate regression analyses highlighting significant predictors of diagnostic delays

Variable	Delta (years)	P-value	N
Demographic details			
Age: >50 years (*vs* age ≤50 years)	–0.3	0.8	189
Gender: female (*vs* male)	0.6	0.6	172
Race: White (*vs* non-White)	0.54	0.92	277
Ethnicity: Hispanic/Latino (*vs* non-Hispanic/Latino)	–12.26	0.02	13
ADI: IQR (*vs* LQ)	8.2	0.03	105
ADI: UQ (*vs* LQ)	–0.57	0.71	50
Driving distance from Mayo: IQR (*vs* LQ)	0.55	0.76	146
Driving distance from Mayo: UQ (*vs* LQ)	–5.1	0.02	73
Symptom history[a]			
Nasal/throat symptoms	0.05	0.97	78
Musculoskeletal symptoms	–1.37	0.09	113
Pulmonary symptoms	–4.1	0.02	71
Cardiovascular symptoms	–1.25	0.57	33
Dermatologic symptoms	–2.08	0.13	124
Anaphylaxis	–3.65	0.13	38
Fatigue	–1.09	0.50	91
Lightheadedness	–1.06	0.54	66
Abdominal symptoms	–1.02	0.47	118
Neuropsychiatric symptoms)	2.23	0.14	110

[a]All symptom history comparisons are *versus* the group of patients without a history of the corresponding symptom.

Predictors of diagnostic delays from multivariate regression analyses

- Multivariate regression analyses were performed to assess predictors of diagnostic delays (**Table 5**)
- Patients of non-Hispanic/Latino ethnicity and those with a driving distance less than 72.1 miles (lower quartile) *versus* more than 352.7 miles (upper quartile) were more likely to experience diagnostic delays
- Patients without pulmonary symptoms were more likely to experience diagnostic delays than those with these symptoms

Conclusions

- We identified a cohort of patients with SM, a clonal disorder of mast cells characterized by non-specific symptoms, and characterized their diagnostic journey
- Real-world studies indicate SM diagnostic delays in the care continuum[8,12]
- The estimated median diagnosis time per LLM was 1.8 years, similar to that seen with a manual approach in this data set (median 1.7 years), but shorter than the median time of 6 years from symptom onset to diagnosis that has been reported previously[8]; this may be due to these patients being evaluated at an SM center of excellence (COE)
- We uncovered non-Hispanic/Latino ethnicity and shorter distance from the clinic as possible predictors of delayed diagnosis time
- Prior to diagnosis, dermatological symptoms were the most frequent, followed by gastrointestinal, musculoskeletal, neuropsychological, and allergic manifestations
- Per the multivariate analysis, the absence of pulmonary symptoms was the most salient predictor of longer diagnostic delays
- The current LLM-based ascertainment of SM-related symptoms is comparable with symptom assessment by humans at an SM COE. Continued refinement of the model may provide an opportunity to further shorten the time to diagnosis beyond provider symptom recognition of SM

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. Brockow K. *Immunol Allergy Clin North Am.* 2014;34:283–295; 5. Cohen SS et al. *Br J Haematol.* 2014;166:521–528; 6. van Doormaal JJ et al. *J Allergy Clin Immunol.* 2013;131:1429–1431.e1421; 7. Bergstrom A et al. *Acta Oncol.* 2024;63:44–50; 8. Mesa RA et al. *Cancer.* 2022;128:3691–3699; 9. Hermine O et al. *PLoS One.* 2008;3:e2266; 10. van Anrooij B et al. *Allergy.* 2016;71:1585–1593; 11. Akin C et al. *J Allergy Clin Immunol.* 2022;149:1912–1918; 12. Yang X et al. *Chin Med J* (Engl). 2025;138:130–142

Acknowledgements

Editorial support under the direction of the authors was provided by Rachel O'Meara, PhD of Prime (Knutsford, UK). Writing and editorial support provided by Prime was funded by Blueprint Medicines Corporation

Conflicts of interest / Disclosures

Dr Yi has received consulting fees from Blueprint Medicines Corporation, research funding from Blueprint Medicines Corporation, and has participated in the speakers' bureau for Curio Science.

For all author disclosures, please contact 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.