Leveraging ChatGPT for Real-World Systemic Lupus Erythematosus Data Curation from Electronic Health Records: A Feasibility Study

Lixia Yao,[1,2] Liwam Gebru,[1] Biruk Aweke,[1] Jay Patel,[1] Ernest Vina,[3] David Fleece,[1] and Huanmei Wu[1]

[1]Temple University, Philadelphia, PA; [2] Polygon Health Analytics LLC, Philadelphia, PA; [3]University of Arizona, Tucson, AZ

BACKGROUND

- Real world data (RWD) can be used to gain valuable insights into disease patterns, treatment outcomes, and healthcare utilization and economics.
- They have also been utilized to identify medical gaps, enhance patient care, and drive innovative therapeutic interventions. However, a significant portion (~80%) of essential clinical data is trapped in unstructured text within clinical notes.

  Our goals were to evaluate:

  - 1) the adequacy of unstructured clinical notes in capturing clinical details (i.e., 22 clinical & immunologic variables) to support meeting the 2019 European League Against Rheumatism and American College of Rheumatology (EULAR/ACR) classification criteria for SLE;
  - 2) the efficacy of the ChatGPT model in identifying those variables outlined in the EULAR/ACR criteria.

METHODS

- We leveraged the Natural Language Processing (NLP) model, ChatGPT 4.0, to analyze clinical notes to identify potential diagnosis of systemic lupus erythematosus (SLE).
- De-identified medical records were obtained from Temple University Health System's EHR system (Epic) which comprised patients >=18 years of age with >=1 rheumatology visit and >=1 ICD-10-CM code of M32.* (SLE) between 1/1/2012 and 5/31/2022.
- Using a Python script, we conducted an automatic search for various mentions of SLE in rheumatology visit notes. We then identified those with a positive anti-nuclear antibody (ANA) test.
- A thorough manual review of the rheumatology notes were conducted by trained medical professionals to determine if patients met the EULAR/ACR classification criteria.
- We additionally employed ChatGPT 4.0 to extract the EULAR/ACR clinical variables.

RESULTS

- The initial Python search yielded 16,124 rheumatology notes from 445 unique patients. Among these, 1,147 rheumatology notes from
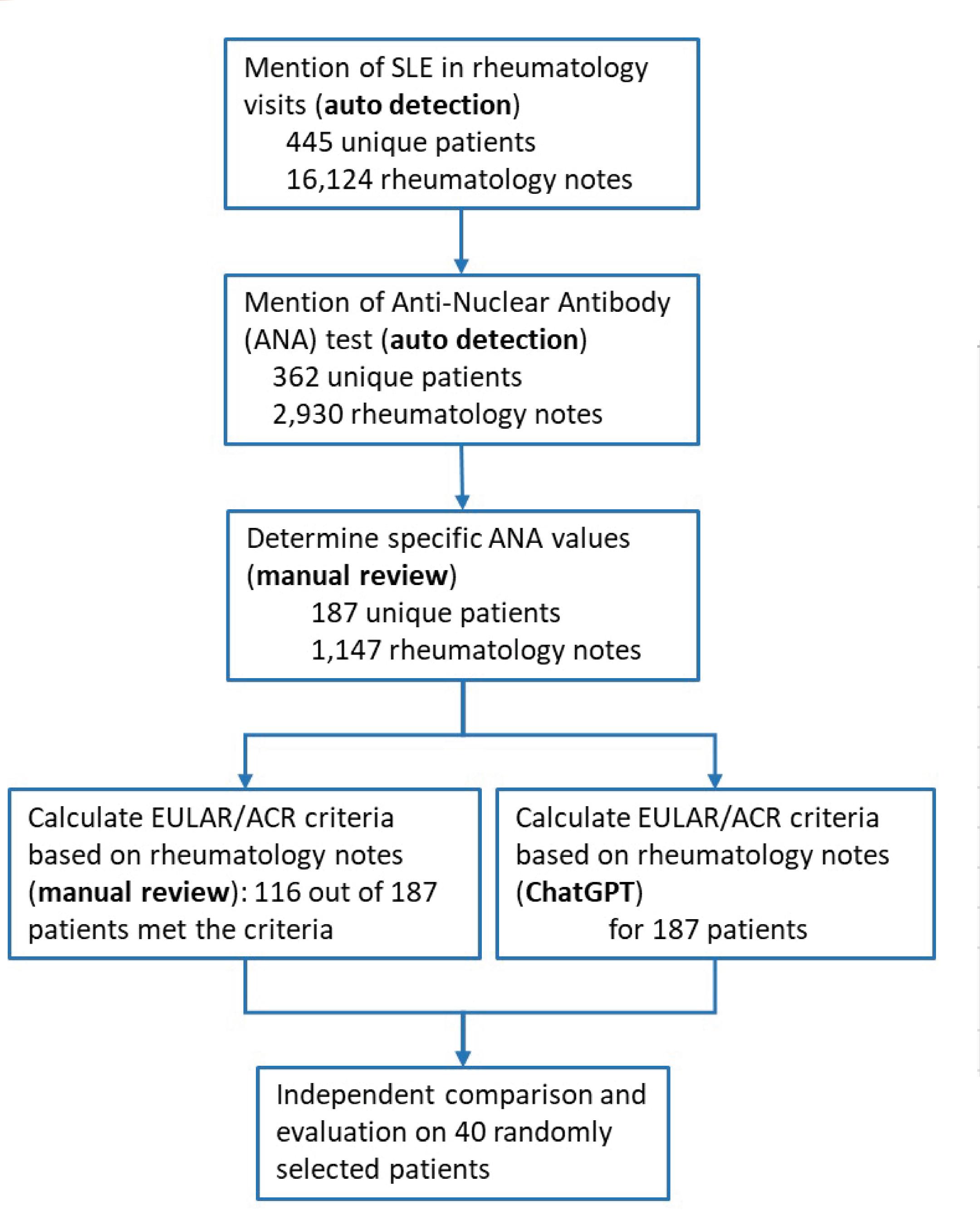

Figure 1: Overall study workflow

187 unique patients had a positive ANA (with values).
- The manual review revealed that 116 of the 187 patients met the EULAR/ACR criteria.
- There was a remarkably high level of agreement between the assessments made by ChatGPT 4.0 and those made by medical professionals: 830/880 clinical & immunologic variables across

40 randomly selected patients.
- Among the 50 annotations of disagreement (Table 1),ChatGPT's interpretations were found to be accurate in 29 cases.
- In some cases, the differentiation between acute cutaneous lupus vs. subacute cutaneous lupus was not clearly documented in the visit notes. In other cases, ChatGPT and human annotators had to infer the classification based on patient-reported symptoms.

Clinical variables included in EULAR/ACR classification of SLE	Discrepancy between medical professional and chatGPT	Rightness by chatGPT	Rightness by medical professional
Alopecia	4	3	1
Oral ulcer	1	0	1
Subacute cutaneous or discoid lupus	8	3	5
Acute cutaneous lupus	5	3	2
Pleural or pericardial effusion	2	1	1
Joint involvement	12	10	2
Proteinuria	4	4	0
Antiphospholipid antibodies	4	2	2
Low C3 or low C4	1	1	0
Low C3 and low C4	3	1	2
anti-dsDNA antibody/anti-smith antibody	6	1	5
Total	50	29	21

Table 1: Analysis of disagreements of annotations by medical professionals and ChatGPT

CONCLUSIONS

- Our findings confirm that unstructured clinical notes contain sufficient clinical details into SLE patient care, including 22 clinical & immunologic variables specified in an established classification criterion for SLE.
- The utilization of state-of-the-art NLP techniques has significant promise in identifying those with SLE, potentially enhancing existing SLE research using RWD.