Natural language processing of clinical notes enables early inborn error of immunity risk ascertainment

Kirk Roberts, PhD,[a] Aaron T. Chin, MD,[b] Klaus Loewy, MS,[c] Lisa Pompeii, PhD,[d] Harold Shin, MS,[e] and Nicholas L. Rider, DO[f,g] *Houston, Tex; Los Angeles, Calif; Cincinnati, Ohio; and Lynchburg and Roanoke, Va*

Background: There are now approximately 450 discrete inborn errors of immunity (IEI) described; however, diagnostic rates remain suboptimal. Use of structured health record data has proven useful for patient detection but may be augmented by natural language processing (NLP). Here we present a machine learning model that can distinguish patients from controls significantly in advance of ultimate diagnosis date.
Objective: We sought to create an NLP machine learning algorithm that could identify IEI patients early during the disease course and shorten the diagnostic odyssey.
Methods: Our approach involved extracting a large corpus of IEI patient clinical-note text from a major referral center's electronic health record (EHR) system and a matched control corpus for comparison. We built text classifiers with simple machine learning methods and trained them on progressively longer time epochs before date of diagnosis.
Results: The top performing NLP algorithm effectively distinguished cases from controls robustly 36 months before ultimate clinical diagnosis (area under precision recall curve > 0.95). Corpus analysis demonstrated that statistically enriched, IEI-relevant terms were evident 24+ months before diagnosis, validating that clinical notes can provide a signal for early prediction of IEI.
Conclusion: Mining EHR notes with NLP holds promise for improving early IEI patient detection. (J Allergy Clin Immunol Global 2024;3:100224.)

Key words: *Natural language processing, machine learning, text mining, inborn errors of immunity, primary immunodeficiency, diagnosis, artificial intelligence*

Abbreviations used

AUPRC:	Area under precision recall curve
EHR:	Electronic health record
ICD:	International Classification of Disease
IEI:	Inborn errors of immunity
ML:	Machine learning
NLP:	Natural language processing
UMLS:	Unified Medical Language System

Individuals with rare diseases experience lengthy diagnostic odysseys, increased morbidity, and adverse health outcomes at disproportionate rates.[1-3] Similarly, persons living with undiagnosed primary immune disorders/inborn errors of immunity (IEI) often come to diagnosis late with resultant suboptimal health status.[4,5] Reasons for delayed diagnosis among individuals with IEI include lack of IEI sign and symptom recognition by health care providers, clinical heterogeneity masking the primary disorder, and suboptimal use of health system science to detect patients within populations.[6-8] While education campaigns remain critical for raising awareness about IEI, leveraging health system data has shown promise for predicting immunologic risk across populations.[8-11] Machine learning (ML) and other computational approaches are emerging to enable earlier and more accurate diagnoses.[8,12,13] To date, application of ML and artificial intelligence has been limited to analyzing structured electronic health record (EHR), laboratory, and genomic data; however, use of unstructured data such as free text remains largely untapped as a resource.[8,10,12,14]

Natural language processing (NLP) refers to the branch of computer science and artificial intelligence whereby free text is represented in a manner that renders it interpretable by computers. NLP has been widely applied to biomedicine and used in pursuit of optimizing diagnoses from clinical notes.[15-17] Recent work in this arena has also called for a focus on leveraging EHR text to more precisely extract and understand symptoms for various disease states.[18] Given strides that the NLP research community has made toward improving diagnostic rates for persons with conditions such as Alzheimer disease, subclinical strokes, thromboembolic disease, and childhood asthma, there is optimism for its utility in rare diseases such as IEI.[19-22]

In contrast to structured EHR data, unstructured free text permits great flexibility in modeling various features of relevance for predicting and understanding IEI from the written language of health care providers. Natural language is not limited by precoordinated terms inherent in ontologies like the International Classification of Disease (ICD)-10. For example, relevant clinical features that could herald undiagnosed IEI include recurrent

From [a]the McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston; [b]the Division of Immunology, Allergy, and Rheumatology, University of California, Los Angeles; [c]the Texas Children's Hospital, Houston; [d]the Department of Patient Services, Cincinnati Children's Hospital Medical Center, Cincinnati; [e]the College of Osteopathic Medicine, Liberty University, Lynchburg; [f]the Division of Health System & Implementation Science, Virginia Tech Carilion School of Medicine, and [g]the Section of Allergy and Immunology, Carilion Clinic, Roanoke.

Received for publication November 6, 2023; revised December 24, 2023; accepted for publication January 21, 2024.
Available online February 2, 2024.

Corresponding author: Nicholas L. Rider, DO, Virginia Tech Carilion School of Medicine, 1 Riverside Circle, 249 Roanoke, VA 24016. E-mail: nick70@vt.edu.

The CrossMark symbol notifies online readers when updates have been made to the article such as errata or minor corrections
2772-8293

https://doi.org/10.1016/j.jacig.2024.100224

fevers and recurrent infections of varying type. These clinical concepts can be described in complex ways via written text which allow for more rich assessments about event frequency unlike structured codes that are only binary (eg, present or absent). For example, text-derived descriptors of recurrent fever include "patient has recurrent fevers," "patient has had fevers without origin for 2 weeks," and "patient's fever reached 103°F for 10 days in a row." This free-text information may relate to an elicited history from a parent/caregiver in the outpatient setting and/or they may be descriptions of an observed clinical course in a hospital setting. In comparison, structured codes would only provide insight into frequency of fevers as a count of individual fever codes (eg, 780.86) or if the health care provider enters a more precise code indicating "relapsing fever" (eg, A68.9). In either case, the richness of concern about "recurrent fever" is often underappreciated with structured data.

The purpose of this study was to develop an NLP system capable of early IEI patient recognition using real-world clinical notes. Our experimental goal was to quantitatively assess NLP classifier capacity for IEI patient diagnosis before their receipt of an IEI diagnosis via clinical practice. In this report, we describe an NLP system that we trained on coded IEI cases and similarly matched controls. Case verification was undertaken via clinical immunology domain expertise and the computational approach via data extraction from real-world EHR free text. We found our system to be capable of early IEI diagnosis using only clinical-note free text among a cohort of verified IEI subjects.

METHODS

Data description

Deidentified EHR information was extracted from the Texas Children's Hospital Epic system on 5,901 verified IEI patients and 285,614 controls. Inclusion criteria for IEI subjects required that an individual had at least 2 primary immune disorder/IEI-representative diagnostic codes entered in their health record on different dates; controls were specified by the complete absence of any such codes.[23] To establish timing of clinical diagnosis for each IEI subject, a time stamp flag was placed at the date/time for which the first specific IEI code was entered.

Creation of the control set involved matching with IEI cases according to sex, ethnicity/race, age, and number of clinical notes (see Appendix A in the Online Repository available at www.jaci-global.org). Controlling for clinical-note count was critical for proper control matching because IEI patients tended to have far greater interaction with the health care system, thereby representing a potential source of bias. Accounting for this finding reduced the likelihood that an ML classifier would simply learn to identify patients with large quantities of data as opposed to learning the underlying early indicators of IEI. To account for potential imbalanced evaluations, 5 controls were matched for every IEI patient, which resulted in 5,901 IEI cases (see Appendix B in the Online Repository) and 28,100 controls.

For each patient, all clinical notes (eg, progress, history and physical, consultation, emergency care, discharge summary, and nursing) were sorted by note-creation time stamp, tokenized, and processed by MetaMap for medical concept extraction, which was based on the Unified Medical Language System (UMLS).[24] Because we were interested only in the value of the clinical notes themselves, other patient data (eg, demographics, diagnostic codes, laboratory test results) were not considered in this study.

Manual verification

We manually reviewed the entire text record for a subset of our IEI cohort (50 patients) to ensure that no mention of a primary immune/IEI disorder diagnosis was evident before the diagnosis time stamp. From this manual review, we also assessed whether the patient had sufficient information to suggest IEI. We then used this process to systematically remove subjects and diagnoses that were not clearly related to IEI, and retrained and validated our algorithms (see Fig E1 in the Online Repository available at www.jaci-global.org). Additionally, we manually reviewed the records for a subset of our control cohort (50 patients) to ensure that IEI patients were excluded.

Experimental setup

Our principal aim was to assess the ability for the NLP classifier to identify IEI patients before their diagnosis in normal clinical practice. As a result, patient data were truncated to simulate the data available at varying prediagnosis time points. We experimented with different month cutoffs, starting with 0 months before diagnosis (ie, all notes were available to a clinician at the time of diagnosis for a given patient) and then proceeding backward from the diagnosis date in 3-month increments (Fig 1).

Given the low real-world prevalence of IEI as well as our inability to simulate this case-to-control imbalance, we focused on standard 1:1 case–control experimentation and relied on metrics that were more invariant to case balancing. In all the experiments we describe, the case–control matching procedure was repeated and optimized for each ML experiment. For these reasons, a 1:5 case–control balance was used to ensure that sufficient control data were available to match case data for each time epoch.

ML model

Because of the large amount of study data, we were unable to use recommended state-of-the-art deep learning methods such as bidirectional encoder representations from transformers (aka BERT), T5, or GPT.[25-27] While methods have been proposed to handle longer clinical sequences (eg, Si and Roberts[28] propose a 3-tiered BERT model, whereas Li et al[29] propose Clinical-Longformer and Clinical-BigBird models), our text data length was significantly larger than what was evaluated in those tasks. Therefore, we focused on a relatively simple model that allowed for the assessment of the potential of NLP classification of IEI without the significant engineering hurtles needed to scale a deep learning model to the size of the data used in our experiments. Specifically, we utilized a linear support vector machine model with L2-regularized L2-loss dual form with a C value of 0.5. We experimented with different feature sets: cased and uncased unigrams (1-token sequences), cased and uncased bigrams (2-token sequences), and UMLS concepts extracted by MetaMap (represented by the concept unique identifier). Model performance was assessed by calculating precision, recall, F1 score, area under the receiver operator curve, and area under precision recall curve (AUPRC) on the test set. Model prediction stability over time was assessed by comparing the AUPRC for models trained and tested at various intervals from ultimate diagnosis. Table I lists the relevant data science terms.

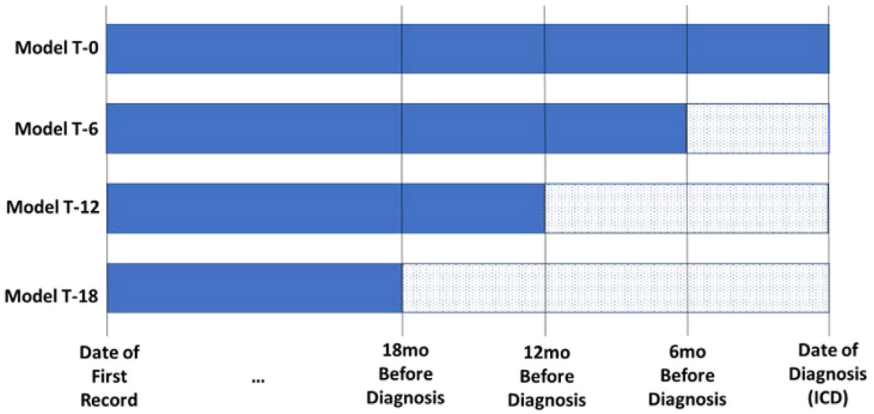

FIG 1. Temporal model configurations. *Blue* indicates subset of patient notes considered by each model. For example, model T-6 considers all notes that occurred at least 6 months before IEI diagnosis.

TABLE I. Relevant data science terms

Term	Definition
Bigram	Any consecutive 2-word pair (eg, "became ill," "was sick").
Corpus	Body of text from which NLP systems are trained and tested.
MetaMap	National Library of Medicine tool developed for recognizing UMLS text concepts.
Support vector machine	Supervised linear ML model type that mathematically seeks to model boundaries between classes of data.
Token	Piece of text taken from corpus and made identifiable for purposes of NLP.
Unified Medical Language System (UMLS)	Set of files and software used to assemble important biomedical and health-relevant vocabularies and standards.
Unigram	Any single text item or word.

RESULTS

Table II displays our cohort demographic information, including the prediagnosis subset of notes for IEI patients, yielding a total of 5,901 IEI cases and 28,100 controls. Of 5,901 IEI cases, 5,207 (88%) contained clinical-note text before an IEI diagnosis. The most prevalent IEI subcategories fell within ICD-10 D80.1 (Hypogammaglobulinemia; n = 1904, 32%), D82.1 (DiGeorge Syndrome; n = 869, 15%), D80.2 (Selective IgA Deficiency; n = 775, 13%), and D81.9 (Combined Immunodeficiency; n = 302, 5%); all subcategories and frequencies are noted in Appendix B in the Online Repository. Our matching ensured demographic similarities among cases and controls with a distribution representative of the population that the hospital served. We also note a generally sex-balanced cohort, with only a slight male predominance.

From our original manual review of IEI subject text before diagnosis, we noted very few (~2%) explicit mentions of IEI before diagnosis. Most subjects had verifiable clinical features to suggest immune dysfunction (~82%), as read by an expert clinical immunologist. Removal of inappropriate diagnoses (eg, B_{12}-deficiency spectrum disease) did not appreciably alter the performance of our classifiers. In addition, no control subjects had clearly evident secondary or primary immune disorders on manual chart review.

The results in Table III demonstrate that simple unigram features generally outperform other basic features (uncased unigrams, bigrams, uncased bigrams, and UMLS concepts) within the setting where all prediagnosis notes are utilized. The bigram features had slightly higher recall (89.27 vs. 88.27 for unigrams) and F1 (89.73 vs 89.47); however, unigrams performed better in all other metrics, including AUPRC.

The experimental findings using shifting temporal windows (Fig 1) are presented in Table IV. We noted an impressive minimal drop in NLP classifier performance as the diagnostic window is shifted backward in time (AUPRC = 0.94-0.95). As we expected, as the prediagnosis window shifts backward in time, the number of patients with notes available was reduced. For example, of the total 5,205 patients with at least 1 prediagnosis note, only 3,883 (75%) have a note dated at least a month before diagnosis.

On text/corpus review, some of the most statistically enriched terms (listed in the Online Repository) among IEI subjects included "hypogammaglobulinemia," "leukopenia," "neutropenia," and "chronic sinusitis" with other expected concepts. We also noted an element of noise in addition to several EHR artifacts on UMLS and term review. For example, EHR artifacts included terms related to institutional vendor transition, such as "import" and "migrated." We also noted some element of MetaMap concept noise owing to its overeager matching for the word "was" being mapped to the concepts "Wiskott-Aldrich syndrome" and "*WAS* gene." These findings were overbalanced by a substantial number of clinically relevant concepts and terms relating to IEI and clinical immunology, as shown in the Online Repository.

DISCUSSION

In this report, we showcase an NLP system trained on coded IEI cases and similarly matched controls. The case verification was

TABLE II. Patient demographics both overall and before diagnosis

Characteristic	Controls	Total IEI cases	Prediagnosis IEI cases
No. of patients	28,100	5,901	5,207
No. of notes per patient			
Minimum	1	1	1
Maximum	8,302	7,543	5,585
Mean	312.2	378.2	170.9
Median	95	115	32
No. of tokens per patient			
Minimum	1	2	2
Maximum	7,891,726	11,711,001	4,637,738
Mean	209,296.9	299,310.6	121,504.7
Median	53,049	77,249	17,935
Ethnicity/race			
White Non-Hispanic	14,290 (50.9%)	2,990 (50.7%)	2,656 (51.0%)
Hispanic	7,755 (27.6%)	1,688 (28.6%)	1,486 (28.5%)
Non-Hispanic	2,763 (9.8%)	569 (9.6%)	475 (9.1%)
Asian	1,112 (4.0%)	222 (3.8%)	205 (3.9%)
American Indian, Hawaiian, multiracial, other	419 (1.5%)	89 (1.5%)	87 (1.7%)
Unknown	1,761 (6.3%)	343 (5.8%)	296 (5.7%)
Sex			
Female	13,450 (47.9%)	2,728 (46.2%)	2,416 (46.4%)
Male	14,650 (52.1%)	3,173 (53.8%)	2,789 (53.6%)

TABLE III. Experiments with different classification features using all prediagnosis data and 1:1 case–control matching

Feature set	Precision	Recall	F1	AUROC	AUPRC
Unigrams	90.71*	88.27	89.47	0.96073*	0.96529*
Unigrams (uncased)	89.33	86.92	88.11	0.95416	0.95484
Bigrams	90.43	89.04*	89.73*	0.95803	0.96108
Bigrams (uncased)	90.59	88.85	89.71	0.95893	0.96166
UMLS concepts	87.48	88.65	88.06	0.95173	0.95470

AUROC, Area under receiver operator curve.
*Highest performance for that particular metric.

TABLE IV. Experiments using increasing amounts of prediagnosis censoring with unigrams used as features

Prediagnosis month	No. of PID patients	Precision	Recall	F1	AUROC	AUPRC
0	5,205	90.71	88.27	89.47	0.96073	0.96529
1	3,883	89.57	86.34	87.93	0.95312	0.95504
2	3,690	87.80	89.70	88.74	0.95556	0.95629
3	3,532	89.02	82.72	85.76	0.94079	0.94441
6	3,186	89.46	88.05	88.75	0.96265	0.96619
9	2,959	89.08	88.47	88.78	0.96463	0.96970
12	2,757	89.96	88.00	88.97	0.95935	0.96210
15	2,560	87.95	85.55	86.73	0.93701	0.93061
18	2,403	87.73	80.42	83.91	0.93969	0.94347
21	2,280	92.75	84.21	88.28	0.94868	0.95000
24	2,145	89.22	85.05	87.08	0.95163	0.95269
27	2,017	89.13	81.59	85.19	0.94490	0.95241
30	1,890	90.59	81.48	85.79	0.93900	0.94273
33	1,784	87.43	85.96	86.69	0.94868	0.95529
36	1,681	89.10	82.74	85.80	0.94668	0.95258

AUROC, Area under receiver operator curve; *PID,* primary immune disorder.

undertaken via clinical immunology domain expertise and the computational approach via data extraction from real-world EHR free text. Our system is capable of early IEI diagnosis using only clinical-note free text compared to time of ultimate diagnosis among a cohort of subjects with verified IEI.

With this work, we present the first ML approach for early diagnosis of IEI via NLP. Thus, our approach is novel in scope as well as methodology: we investigate performance across a sliding diagnostic window. Our system shows temporal fidelity, with disease recognition 36 months before clinical IEI diagnosis. This is highly relevant because clinicians seek to make early and precise diagnoses in persons with IEI, but to date, such prediction has not been demonstrated by any artificial intelligence system.

In previous work, we have shown the utility of using structured data for identifying IEI risk (eg diagnosis codes, problem lists), and we expect those efforts to synergize well with this NLP approach. Toward a multimodal approach for IEI patient finding, use of the written natural language leverages a much richer information source and is complementary with other vetted systems. In this way, we bring successes in enabling diagnosis for other clinical domains to the space of clinical immunology.[30-32] We also allow for leveraging clinical-note text written in the language of health care providers in the ways that they describe clinical concepts of relevance for IEI. For example, capturing relevant clinical concepts such as "recurrent pneumonia" or "unusual infection" is not easy via structured ontologies such as ICD codes. However, these concepts can be readily identified by text analysis and are useful for discriminating between common conditions and those that might herald an IEI. Our findings suggest that sufficient information exists within clinical notes to make an early diagnosis, meaning that disease can be detected by a prediction model long before a diagnosis would be made using current clinical workflows.

Our NLP system has the capacity to use 1- or 2-word features extracted from real-world IEI subjects before diagnosis (Table II). Despite model and feature simplicity, the performance is remarkably high and allows for sufficiently robust performance by our classifier across the 36 months leading up to diagnosis (Fig 1). Table IV shows generally stable NLP classifier performance across all metrics over the time epochs studied. Of note, precision is consistently higher than recall, which may reflect real-world heterogeneity of the IEI patient diagnostic odyssey. Our work shows that many IEI patients have evident signs heralding their disease in their clinicians' notes, while others have no such signal and are indistinguishable from controls. This observation may reflect differing signal strengths from patient to patient and from one IEI subtype to another, along with other factors. However, because we used a large and diverse IEI cohort that was controlled for age, sex, and number of clinical notes (a proxy for health system engagement), we expect that the performance metrics will be sufficiently stable in other health systems. These factors were important considerations for training our model and are reflected in the immunologic diversity shown in Appendix B in the Online Repository.

It is important to note that our approach focused on training an NLP system on EHR text from patients with verified IEI *before* their diagnosis was made. This allowed us to use clinician descriptions about IEI subjects during very early stages of clinical evaluations, when the underlying condition was not known. The observation that many IEI diseases were mentioned well before diagnosis suggests that IEI patients were experiencing immune-related symptoms years before their official diagnosis. This finding also motivates the potential for both more advanced NLP information extraction, which aims to remove false positive concept matches, and feature selection methods that (semi)automatically identify IEI-related terms and use those exclusively for model features. We expect that these and other descriptors can be used as heralding elements for IEI, which may be broadly applicable across health care institutions and geographic regions. Through this and future work, we have probed and will continue to probe the skilled narratives clinicians create as they work to understand the nebulous, unclear trajectories of the disease of patients when they initially seek care for IEI. Ultimately, we are building systems capable of augmenting human intuition and ascribing risk within clinical environments.

We expect that this initial proof of concept will lead to improved outcomes for patients with IEI. As we carefully build computational systems that precisely synthesize multimodal EHR data for improving diagnosis and outcomes, we will improve diagnostic sensitivity, thereby mitigating delayed diagnosis among persons with IEI. Given that diagnostic delays remain a major obstacle leading to suboptimal outcomes and that early diagnosis is linked to reduced morbidity and more precise therapies, these efforts advance the field in important ways.[4,33,34]

Our work has some notable limitations. First, our NLP system was constructed from clinical-note text taken from a single US center that uses the English language. Thus, performance in other centers or where other languages are used is unknown and will require model validation and tuning. In addition, we did not augment text inputs with other structured or genomic data that could improve performance broadly. Last, our approach involves only 1:1 case–control matching, which was essential for modeling a rare disease but which is not reflective of real-world clinical encounters.

The next steps in our work will include advanced model development, such as the aforementioned transformer-based models, focusing on approaches that could be more predictably generalizable and have less bias. Experiments are needed to assess the precision and recall/sensitivity of the model when deployed en masse, which will require and include substantially more control patients to evaluate. We also plan to incorporate structured EHR data elements into a multimodal classification system for more comprehensive utilization of all portions of a patient's clinical record. Last, our prior work in building probabilistic models will inform future efforts relating to causal investigations about how and why certain clinical events are associated with one another within distinct IEI disorders.[11]

DISCLOSURE STATEMENT

Research reported in this work was supported by Takeda Pharmaceuticals (K.R., N.L.R.) as well as the National Institute of Allergy and Infectious Diseases under award R21AI164100 (N.L.R., K.R.).

Disclosure of potential conflict of interest: N. L. Rider receives royalties from Wolters Kluwer/UpToDate and is paid consulting fees by Takeda Pharmaceuticals and Pharming Healthcare. The rest of the authors declare that they have no relevant conflicts of interest.

We thank Colin Anderson-Smits, PhD, for critical review of the design and the report.

Clinical implication: Use of structured and unstructured health record data together may greatly improve diagnostic rates for undiagnosed IEI patients.

REFERENCES

1. Isono M, Kokado M, Kato K. Why does it take so long for rare disease patients to get an accurate diagnosis?—A qualitative investigation of patient experiences of hereditary angioedema. PLoS One 2022;17:e0265847.
2. Johansen MB, Bendstrup E, Davidsen JR, Shaker SB, Martin HM. The diagnostic trajectories of Danish patients with autoimmune rheumatologic disease associated interstitial lung disease: an interview-based study. Eur Clin Respir J 2023;10:2178601.
3. Merker VL, Slobogean B, Jordan JT, Langmead S, Meterko M, Charns MP, et al. Understanding barriers to diagnosis in a rare, genetic disease: delays and errors in diagnosing schwannomatosis. Am J Med Genet A 2022;188:2672-83.
4. Branch A, Modi B, Bahrani B, Hildebrand KJ, Cameron SB, Junker AK, et al. Diverse clinical features and diagnostic delay in monogenic inborn errors of immunity: a call for access to genetic testing. Pediatr Allergy Immunol 2021;32:1796-803.
5. Odnoletkova I, Kindle G, Quinti I, Grimbacher B, Knerr V, Gathmann B, et al. The burden of common variable immunodeficiency disorders: a retrospective analysis of the European Society for Immunodeficiency (ESID) registry data. Orphanet J Rare Dis 2018;13:201.
6. Shamriz O, Zahalka N, Simon AJ, Lev A, Barel O, Mor N, et al. *GATA2* deficiency in adult life is characterized by phenotypic diversity and delayed diagnosis. Front Immunol 2022;13:886117.
7. El-Sayed ZA, El-Ghoneimy DH, Ortega-Martell JA, Radwan N, Aldave JC, Al-Herz W, et al. Allergic manifestations of inborn errors of immunity and their impact on the diagnosis: a worldwide study. World Allergy Organ J 2022;15:100657.
8. Rider NL, Coffey M, Kurian A, Quinn J, Orange JS, Modell V, et al. A validated artificial intelligence–based pipeline for population-wide primary immunodeficiency screening. J Allergy Clin Immunol 2023;151:272-9.

9. Quinn J, Modell V, Orange JS, Modell F. Growth in diagnosis and treatment of primary immunodeficiency within the global Jeffrey Modell Centers Network. Allergy Asthma Clin Immunol 2022;18:19.
10. Rider NL, Miao D, Dodds M, Modell V, Modell F, Quinn J, et al. Calculation of a primary immunodeficiency "risk vital sign" via population-wide analysis of claims data to aid in clinical decision support. Front Pediatr 2019;7:70.
11. Rider NL, Cahill G, Motazedi T, Wei L, Kurian A, Noroski LM, et al. PI Prob: a risk prediction and clinical guidance system for evaluating patients with recurrent infections. PLoS One 2021;16:e0237285.
12. Mayampurath A, Ajith A, Anderson-Smits C, Chang SC, Brouwer E, Johnson J, et al. Early diagnosis of primary immunodeficiency disease using clinical data and machine learning. J Allergy Clin Immunol Pract 2022;10:3002-7.e5.
13. Fang M, Su Z, Abolhassani H, Itan Y, Jin X, Hammarström L. VIPPID: a gene-specific single nucleotide variant pathogenicity prediction tool for primary immunodeficiency diseases. Brief Bioinform 2022;23:bbac176.
14. Emmaneel A, Bogaert DJ, Van Gassen S, Tavernier SJ, Dullaers M, Haerynck F, et al. A computational pipeline for the diagnosis of CVID patients. Front Immunol 2019;10:2009.
15. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 2019;25:433-8.
16. Shah AD, Bailey E, Williams T, Denaxas S, Dobson R, Hemingway H. Natural language processing for disease phenotyping in UK primary care records for research: a pilot study in myocardial infarction and death. J Biomed Semantics 2019; 10(suppl 1):20.
17. Ambrosy AP, Parikh RV, Sung SH, Narayanan A, Masson R, Lam PQ, et al. A natural language processing–based approach for identifying hospitalizations for worsening heart failure within an integrated health care delivery system. JAMA Netw Open 2021;4:e2135152.
18. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. J Am Med Inform Assoc 2019;26:364-79.
19. Liu N, Luo K, Yuan Z, Chen Y. A transfer learning method for detecting Alzheimer's disease based on speech and natural language processing. Front Public Health 2022;10:772592.
20. Kent DM, Leung LY, Zhou Y, Luetmer PH, Kallmes DF, Nelson J, et al. Association of silent cerebrovascular disease identified using natural language processing and future ischemic stroke. Neurology 2021;97:e1313-21.
21. Shi J, Hurdle JF, Johnson SA, Ferraro JP, Skarda DE, Finlayson SRG, et al. Natural language processing for the surveillance of postoperative venous thromboembolism. Surgery 2021;170:1175-82.
22. Seol HY, Rolfes MC, Chung W, Sohn S, Ryu E, Park MA, et al. Expert artificial intelligence–based natural language processing characterises childhood asthma. BMJ Open Respir Res 2020;7:e000524.
23. American Academy of Allergy, Asthma & Immunology (AAAAI). ICD-10 codes for immunodeficiency. March 2017. Available at: https://www.aaaai.org/Aaaai/media/MediaLibrary/PDF%20Documents/Practice%20Management/finances-coding/ICD-10-Codes-Immunodeficiencies.pdf.
24. Demner-Fushman D, Rogers WJ, Aronson AR. MetaMap Lite: an evaluation of a new Java implementation of MetaMap. J Am Med Inform Assoc 2017;24:841-4.
25. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: human language technologies, volume 1 (long and short papers). Minneapolis (Minn): Association for Computational Linguistics; 2019. p. 4171-86. Available at: https://aclanthology.org/N19-1423.
26. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. J Mach Learn Res 2020;21:140:5485-551.
27. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. In: Advances in neural information processing systems. NurIPS 2020. Red Hook (NY): Curran Associates; 2020. p. 1877-901. Available at: https://proceedings.neurips.cc/paper_files/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.
28. Si Y, Roberts K. Three-level hierarchical transformer networks for long-sequence and multiple clinical documents classification. arXiv, April 27, 3032, updated December 16, 2021. https://doi.org/10.48550/arXiv.2104.08444.
29. Li Y, Wehbe RM, Ahmad FS, Wang H, Luo Y. Clinical-Longformer and Clinical-BigBird: transformers for long clinical sequences. arXiv, January 26, 2022, updated April 15, 2022. https://doi.org/10.48550/arXiv.2201.11838.
30. Mermin-Bunnell K, Zhu Y, Hornback A, Damhorst G, Walker T, Robichaux C, et al. Use of natural language processing of patient-initiated electronic health record messages to identify patients with COVID-19 infection. JAMA Netw Open 2023;6:e2322299.
31. Soni A, Amrhein B, Baucum M, Paek EJ, Khojandi A. Using verb fluency, natural language processing, and machine learning to detect Alzheimer's disease. Annu Int Conf IEEE Eng Med Biol Soc 2021;2021:2282-5.
32. Berman AN, Ginder C, Sporn ZA, Tanguturi V, Hidrue MK, Shirkey LB, et al. Natural language processing for the ascertainment and phenotyping of left ventricular hypertrophy and hypertrophic cardiomyopathy on echocardiogram reports. Am J Cardiol 2023;206:247-53.
33. Carrabba M, Dellepiane RM, Cortesi M, Baselli LA, Soresina A, Cirillo E, et al. Long term longitudinal follow-up of an AD-HIES cohort: the impact of early diagnosis and enrollment to IPINet centers on the natural history of Job's syndrome. Allergy Asthma Clin Immunol 2023;19:32.
34. Boz V, Tesser A, Girardelli M, Burlo F, Pin A, Severini GM, et al. Inborn errors of immunity in children with autoimmune and allergic complaints: a single center experience from diagnosis to treatment. Biomedicines 2023;11:1299.