Title

Proceedings from the inaugural Artificial Intelligence in Primary Immune Deficiencies (AIPID) conference.

Journal

Journal of Allergy and Clinical Immunology, 153(3)

Authors

Rivière, Jacques
Soler Palacín, Pere
Butte, Manish

Publication Date

2024-03-01

DOI

10.1016/j.jaci.2024.01.002

Peer reviewed

Published in final edited form as:

J Allergy Clin Immunol. 2024 March ; 153(3): 637–642. doi:10.1016/j.jaci.2024.01.002.

Proceedings from the inaugural Artificial Intelligence in Primary Immune Deficiencies (AIPID) conference

Jacques G. Rivière, MD[a,b,c,d], **Pere Soler Palacín, MD, PhD, MS**[a,b,c,d], **Manish J. Butte, MD, PhD**[e,f,g]

[a]Infection and Immunity in Pediatric Patients Research Group, Vall d'Hebron Barcelona Hospital Campus;

[b]Pediatric Infectious Diseases and Immunodeficiencies Unit, Hospital Infantil i de la Dona, Vall d'Hebron Barcelona Hospital Campus;

[c]Jeffrey Modell Diagnostic and Research Center for Primary Immunodeficiencies, Barcelona;

[d]Universitat Autònoma de Barcelona;

[e]Division of Immunology, Allergy, and Rheumatology, Department of Pediatrics, University of California Los Angeles.

[f]Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles.

[g]Department of Human Genetics, University of California Los Angeles.

Abstract

Here, we summarize the proceedings of the inaugural Artificial Intelligence in Primary Immune Deficiencies conference, during which experts and advocates gathered to advance research into the applications of artificial intelligence (AI), machine learning, and other computational tools in the diagnosis and management of inborn errors of immunity (IEIs). The conference focused on the key themes of expediting IEI diagnoses, challenges in data collection, roles of natural language processing and large language models in interpreting electronic health records, and ethical considerations in implementation. Innovative AI-based tools trained on electronic health records and claims databases have discovered new patterns of warning signs for IEIs, facilitating faster diagnoses and enhancing patient outcomes. Challenges in training AIs persist on account of data limitations, especially in cases of rare diseases, overlapping phenotypes, and biases inherent in current data sets. Furthermore, experts highlighted the significance of ethical considerations, data protection, and the necessity for open science principles. The conference delved into regulatory frameworks, equity in access, and the imperative for collaborative efforts to overcome these obstacles and harness the transformative potential of AI. Concerted efforts to successfully integrate AI into daily clinical immunology practice are still needed.

Corresponding author: Manish J. Butte, MD, PhD, Pediatrics 10833 Le Conte Ave, 12-430 MDCC, Los Angeles, CA 90095. mbutte@mednet.ucla.edu.

DISCLOSURE STATEMENT

Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

Keywords

Artificial intelligence; machine learning; large language models; natural language processing; electronic health records; inborn errors of immunity; diagnosis; ethics

Because of the rapid growth and interest in the fields of artificial intelligence (AI) and machine learning (ML), we convened a conference, the Artificial Intelligence in Primary Immune Deficiencies (AIPD) conference, to discuss applications of these approaches to inborn errors of immunity (IEIs) (also known as primary immunodeficiency diseases) at the Barcelona Supercomputing Center from June 19 to June 21, 2023. The conference was endorsed by the main scientific societies in the field: the European Society for Immunodeficiencies, the Latin American Society for Immunodeficiencies (LASID), and the Clinical Immunology Society, as well as by key stakeholders in the field, such as the Jeffrey Modell Foundation, the International Patient Organization for Primary Immunodeficiencies, and the Immune Deficiency Foundation. The conference brought together patient advocates and diverse research groups primarily from Europe and the Americas. This workshop summary, authored by the hosts of the meeting, aims to offer context and summarize all the presentations, organized by themes. The key topics raised include the use of AI and ML to facilitate the diagnosis of IEIs, the challenges of collecting and using data for research purposes in rare diseases, the use of software called natural language processing (NLP) and large language models (LLMs) to interpret medical notes, and considerations of ethics and implementation. The aim of the conference was to share the latest published and unpublished research, identifying gaps in knowledge and capabilities, highlighting unmet needs, and thus setting a path for potential uses of AI in the field of IEIs.

EXAMPLES OF AI IN FACILITATING THE DIAGNOSIS OF IEIs

As in many other areas of clinical endeavor, AI and ML have the potential to revolutionize the diagnosis of IEIs. Long delays in making diagnoses of IEIs continue to defy the best efforts of clinical immunologists and patients.[1] One challenge has been that no 2 patients with IEIs present the same way. We teach medical students (and AI algorithms) to recognize heart attacks by matching patterns exhibited by patients, such as “crushing substernal chest pain” in patients’ history and elevated troponin levels in their laboratory results, knowing that not every patient with a myocardial infarction will fit this pattern but most will. In the case of IEIs, the challenge arising is that the variability of diagnoses is much broader, ranging from chest pain due to aortic aneurysm (eg, as seen in STAT1 gain of function) to strokes (eg, ADA2 deficiency) to deep fungal abscesses (eg, chronic granulomatous disease) to type 1 diabetes (eg, immunodysregulation polyendocrinopathy enteropathy, X-linked) to cutaneous yeast infections (eg, autoimmune regulator deficiency), and so forth. Some patients will have many episodes and severe clinical features, whereas other patients will lack some clinical features entirely. It is important to also remember that many clinical features of IEIs in our literature may be biased in that they are based primarily on cases in North America and Europe and may neglect presentations from patients living in rural and resource-poor settings and those living with tropical endemic diseases. Dr Condino-Neto from the University of são Paolo and the LASID reported that the LASID disease

registry contains examples of infection phenotypes that differ considerably between Latin American patients and those in Europe or North America (eg, differences in fungal and bacterial infections observed in chronic granulomatous disease[2] or bacillus Calmette-Guérin disease (BCGosis) following vaccination with bacillus Calmette-Guérin).[3] The challenges in diagnosing these conditions owing to their variable and often overlapping clinical features emphasize the need for an individual patient–centered approach to care rather than a traditional diagnosis approach.[4] In that light, how does one successfully teach an AI algorithm to recognize that one person with only recurrent sinus infections and another with only autoimmune hemolytic anemia have the same "diagnosis" (eg, common variable immunodeficiency [CVID])? Furthermore, unlike other rare diseases in which only a few genes underpin the diagnoses, IEIs comprise variants across hundreds of genes, and the vast majority of patients with IEIs still elude genetic diagnoses today. The field of IEIs faces challenges in clarifying how we define our clinical conditions as we move toward integrating AI into our diagnostics.

The motivation to provide early diagnoses for patients with IEIs arises from our desire to shorten the frustrations and reduce unnecessary testing associated with the "diagnostic odyssey."[5–7] Immunoglobulin replacement therapy (IgRT), immunomodulatory treatments, and definitive treatments are expensive; in contrast, avoiding unnecessary treatments and tests saves money. The Modell Foundation has previously reported that early diagnosis of IEIs can save tens of thousands of dollars per patient per year in ICD-diagnosed IEI patients.[8,9] Dr Chris Runken of Health Economics and Outcomes Research at Grifols took a more comprehensive new approach to studying the costs of delayed diagnosis. In unpublished work, he started with the Phar-metrics Plus insurance claims database (2014–2019) and looked at the health care costs associated with almost 600 patients before diagnosis and then 1 to 2 years after diagnosis and treatment with IgRT. His results suggested that early diagnosis may not result in cost savings over 2 years, as making a diagnosis of an immune deficiency often increases expensive doctors' visits and initiation of costly treatments such as IgRT. However, early diagnosis may be financially impactful in terms of longer-term savings through the reduction of complications, especially for children. Additional benefits of early diagnosis include increased work productivity (thanks to fewer missed days of work), increased taxes to society (thanks to increased employment), and increased quality of life.[10] The financial and societal implications of making diagnoses early thus continue to be a strong motivator for the use of AI.

There has been a long history of attempts to facilitate the early diagnosis of IEIs. One of the first efforts to codify the diagnoses that should raise suspicion of immune deficiencies occurred in 1993, with development of the Jeffrey Modell Foundation's 10 Warning Signs. These warning signs highlighted recurrent infections as a major theme, and their broad dissemination resulted in a massive increase in referrals to immunologists.[11] Charlotte Cunningham-Rundles et al were among the first groups to offer calculated refinements to the warning signs as gleaned from large databases of health insurance claims.[12] The Modell Foundation introduced their Software for Primary Immunodeficiency Recognition, Intervention, and Tracking (SPIRIT) analyzer by assigning weights to 350 International Classification of Diseases (ICD) codes that corresponded to the 10 warning signs in a large claims database.[8] These early efforts pioneered the general path that many groups now

undertake with more sophisticated tools (see later). There were also early ML approaches in genomics of IEIs; for example, Jordan Orange et al used support vector machines to uncover a nonpolygenic CVID-like phenotype due to *IRF2BP2*[13] and other CVID genes.[14]

By harvesting data derived from medical records of patients with IEIs, AI and ML algorithms can identify patterns that are indicative of IEIs (Fig 1). This process could be applied to patients whose conditions have not yet been diagnosed, leading to earlier diagnosis and treatment, improving patient outcomes, and reducing costs. Current approaches usually begin by "training" an AI algorithm to recognize IEIs on the basis of clinical features of known cases. This process is much like our current approaches to training medical students, residents, and fellows about IEIs by repeatedly exposing them to actual cases. These "ground truth" cases need to be *bona fide* patients and are not stereotypic or caricatures of IEI diagnoses.

The first step often entails mapping portions of the medical record and standardizing to terms that software can use. Vanderbilt University researchers, including Professor Lisa Bastarache, have established a mapping of billing codes (ICD codes) and a standardized medical curation tool (called Human Phenotype Ontology [HPO]) to convert ICD terms to phecodes.[15] Phecodes are a categorization of diagnoses that is simplified compared with the ICD codes and more efficiently identifies phenotypes for computation. This mapping allows for the representation of "phenotype syndromes" modeled after Mendelian diseases by using clinical phenotypes derived from the electronic health record (EHR). Other groups have used a manual effort to capture phenotypes. Dr Luiza Campos and the INTREPID team at University College London developed a phenotype capture tool that is available in the United Kingdom to collect a list of phenotypes of patients with IEIs. The tool was based on the HPO, and it cataloged the phenotypic features of 886 already-identified patients with IEIs, leading to the recruitment of 600 prospective patients to diagnose CVID.

When an algorithm is being trained, after the various phenotypic features of a patient have been gathered, numeric weights are assigned to each phenotype to indicate their importance in contributing to an IEI diagnosis. For each patient's history in the EHR, their many diagnoses and the corresponding weights are then accumulated to generate a "risk score." Bastarache et al developed an approach called the Phenotype Risk Score (PheRS) that is calculated by summing the weighted clinical characteristics of patients, with each characteristic being weighted by the inverse logarithm of its populational prevalence.[16] The advantages of this approach are its simplicity, legibility, and portability. Dr Bastarache demonstrated the utility of PheRS in that patients with rare genetic diseases have high risk scores, and she has since extended this approach for screening to assess whether patients with high-risk scores in fact carry rare pathogenic genetic variants. Her group performed an early type of this analysis in patients with cystic fibrosis with success.[17] These advances build on foundational software tools that are openly shared: an update of phecodes called PhecodeX,[18] which includes more rare disease phenotypes; an R package (PheRS) that uses phenotype risk scores based on her for rare genetic variants to study Mendelian diseases[19]; and the Phenotype-Genotype reference map (PGRM),[20] a set of genetic associations from hundreds of genome-wide association study–based publications that can be used for high-throughput replication experiments. These approaches have been

used not only to find people experiencing rare diseases but also to unravel novel associations with known diseases.

The process of assigning weights to various diagnoses can be done computationally or by using panels of experts. The PIDCAP project of Jacques G. Rivière from the Vall d'Hebron Barcelona Hospital[21] used an expert-driven model to assign weights to more than 3,500 ICD codes for various diagnoses (eg, pneumonia, diarrhea, cytopenia, etc) and generated risk scores. The initiative was implemented in real-world practice in 2019 in a pilot area encompassing approximately 100,000 individuals. Bronchiectasis, autoimmune disease, cytopenia, and recurrent infections were the most distinguishing warning signs for both adults and children. The system alerts primary care professionals of the risk score by integration into electronic health software. Dr Nicholas Rider of Liberty University introduced his team's research, starting with the Jeffrey Modell Foundation's Software for Primary Immunodeficiency Recognition, Intervention, and Tracking analyzer[22] and moving to more advanced tools.[23–25] He showed that bronchiectasis and splenomegaly were enriched IEIs in the cohort. Certain diagnoses, particularly those related to pneumonias such as *Streptococcus pneumoniae* pneumonia and interstitial pneumonia, demonstrated greater specificity for IEIs. Dr Helen Leavis of Universitair Medisch Centrum Utrech and her team trained an ML approach-based tool using an expert panel to assign weights for 83 diagnosis codes, use of antibiotics, abnormal calculated immunoglobulin levels, and recurrent visit to primary care. The algorithm was then applied to 60,000 people aged 12 to 70 years and used to identify high-risk persons, who were then referred to an immunologist if needed. For those diagnosed by this approach, the estimated cost was around €5,000 per detected case and the estimated diagnostic delay was reduced by 3 years. Professors Manish Butte and Bogdan Pasaniuc from the University of California Los Angeles trained a regression model on approximately 190 patients with CVID at the university and suitable controls.[26] The model used 44 phecodes plus IgG levels to generate a risk score for each patient. This approach made it possible to identify patients with CVID 1 to 4 years before diagnosis. Their approach has extended to the identification of undiagnosed patients across the 5 hospitals of the University of California health system. They found that 60% of undiagnosed patients identified by the algorithm were described in a blinded chart review as "likely to have CVID." Taken together, these new efforts have moved beyond the 10 Warning Signs and can successfully identify patients with IEIs from their EHR. The training data for most efforts still focus on infections, although these efforts have confirmed the importance of looking at autoimmunity and inflammatory features (including constitutional effects such as failure to thrive) as well.

One major difference in existing AI efforts has been the use of claims databases versus individual EHRs to train the algorithms. The former offers advantages of huge numbers of patients (tens of millions), but it lacks data such as laboratory test results and long-term trends (because in the United States, about 20% of individuals change their health insurance every year[27]). The latter are often smaller (covering hundreds of thousands to a few million patients) but offer higher-quality data; longer trends; and additional data in the form of laboratory tests, imaging results, and written notes. Future efforts to feed the voracious appetite for the high-quality data required to properly train AI systems may need to straddle these 2 options.

Moving beyond the well-accepted clinical features of the aforementioned IEIs, Dr Eleonora Gambineri, head of the Immuno-Haem-Onc Unit at Meyer Hospital in Firenze, Italy, introduced new "warning signs" that indicate immune dysregulation and IEIs: cancer, cytopenias (eg, autoimmune), lymphoproliferative disorders, and myelodysplastic disorders.[28] She flipped the conventional thinking that IEIs lead to cancer, instead proposing that cancers could be a red flag for immune defects.

Major challenges in the use of EHRs to diagnose IEIs remain unsolved. First, statistical comparisons for the efficacy of various AI techniques continue to be confounded by the problem stemming from the fact that the "control" subjects being compared in the EHR are actually quite different from the subjects with IEIs. The field needs a cohort of control patients who have "near-miss IEIs," namely, those who have autoimmunity, infections, and inflammatory phenotypes but do not have a monogenic IEI. At this point, assembling such a cohort is difficult if not impossible. Second, overlapping phenotypes with IEIs, such as those that occur in patients with chronic kidney diseases, cirrhosis of the liver, organ transplantation, HIV infection, and cystic fibrosis, require that patients with these diagnoses be excluded from consideration. Third, existing approaches rely on training data that are adult-centric, and accumulating such data requires a number of years of life. Consequently, AI approaches are less effective at identifying children with IEIs. Until new techniques are developed and this bias is mitigated, the likelihood that children will be found by using an EHR-based approach remains low.

LLMs, NLPs, AND TEXT MINING

Because of the low incidence rates, gathering a sufficient amount of data for purposes of research on rare diseases remains a significant challenge. The scarcity of comprehensive data sets notably constrains the advancement of sophisticated data-driven methodologies such as ML and AI, which depend heavily on extensive data sets for training and validation. Pablo González from GMV Innovative Solutions offered 2 different solutions based on harvesting data from collaborative efforts between different institutions through privacy-enhancing technologies, federating learning, and LLMs. Regarding PETs, González highlighted the fact that with the right approaches such as secure multiparty computation, there is no need to choose between data privacy and data usability. Another solution mentioned was to use LLMs to increase the amount and granularity of data from the EHR, organizing loose medical notes into structured concepts such as HPO. He explained that during this process, learning contextual cues from the notes is still an active area of research. For example, many medical notes offer details that require nuance to ascertain the true meaning. Mention of the word *diabetes* in a medical note could indicate that the patient has diabetes, but the words in the sentences around that word could indicate a family history of diabetes instead. When successful, teaching LLMs to read EHR notes with contextual enrichment improves the quality of extracted data and the outcomes of subsequent studies.

Professor Kirk Roberts from the University of Texas Houston and the University of Texas Health School of Biomedical Informatics worked with Dr Rider to develop NLP tools to comprehend medical notes, particularly in the context of identifying individuals with IEIs. He analyzed free text medical records encompassing more than 6,000 patients with IEIs

and compared them with a control group comprising more than 25,000 individuals.[29] An ML model was trained by using the medical notes recorded before the patients' ICD-based diagnosis. Remarkably, the model demonstrated that free text notes offered a powerful capability to predict IEIs in a significant proportion of patients nearly 2 years before formal diagnosis.

Despite the considerable promise of AI, several limitations were acknowledged. One point of concern in training AIs to learn from medical notes on patients with IEIs is in selecting control cases appropriately. Additionally, it is important to harmonize notes across sources, which is an unsolved problem. Finally, there are risks associated with training AIs on notes of inadequate quality. As LLMs are increasingly used in rare diseases to process text, we emphasize the need to ensure data portability and interoperability through the use of federated models.

IMPLEMENTATION, REGULATORY, AND ETHICAL CONSIDERATIONS IN USING AI IN HEALTH RECORDS RESEARCH

Ethical considerations play a major role in the deployment of AI. The ethical challenges in the use of AI systems are many: privacy and data security; biases in the design of an AI model that may lead to discrimination; ethical decision making in the uses of AI; and other concerns related to transparency, accountability, and data governance.

The protection of human subjects and their data requires careful consideration and oversight when conducting AI research on electronic databases. Most health systems today require that research be performed on deidentified databases of health records, which significantly reduces the risk that individual health data could be misused. In the name of data protection, however, health systems often go even further to limit research in EHRs. Unfortunately, draconian policies regarding data protections may carry adverse consequences for patients with IEIs, in which case sharing data may be crucial to further diagnoses. Most surveys of patients with rare diseases consistently show support for data sharing. For example, efforts such as GeneMatcher[30] and Match-Maker Exchange facilitate international collaborations to assemble cohorts of patients with misunderstood genetic variants. These tools are a federated network comprising multiple rare disease databases that match cases based on genetic and health profiles. The databases are deidentified but can link investigators to allow for study of particular patients. Multinational efforts such as Solve-RD leverage databases such as these to organize experts and address unsolved conditions.[31] Rare clinical phenotypes such as IEIs offer unique opportunities compared with those offered by conditions that occur more frequently. Future efforts to advocate for patients with IEIs will require working closely with information management departments of large health systems to ensure that their policies are not overly restrictive for rare disease research. These efforts will have important implications as data policies are applied to AI software. Legal and political scholar Professor Josep Lluis Martí of Pompeu Fabra University, Barcelona, Spain, explained the new European Union Artificial Intelligence Act and its applications to EHR research. This new set of regulations classifies AI systems on the basis of their potential for risk and requires increasing levels of regulation and oversight as the risks increase.

The impact on application of AI to EHR research under these new laws could initially limit efforts as various groups navigate newly established policies and their governance. Furthermore, despite the aforementioned tenet that EHR research should protect the identity of individual patients, there are some situations in which "reidentifying" anonymized data might be necessary (eg, in cases in which a finding requires that urgent treatment be provided).

Besides data protection, another major ethical consideration of using AI in EHR research is related to access, which includes the topics of open science and equity. Electronic health databases are massive repositories of data but often become siloed and not interoperable. Professor Jordi Piera, director of the Digital Health Strategy Office of the Catalan Health Service, and others have reported that these databases can be reworked to increase interoperability (both intravendor and intervendor), as shown in his recent studies.[32,33] One proposed solution includes moving databases over to open source structures in which data are not held hostage by vendors and interoperability is higher. One example is the OpenEHR initiative.[34] Other open tools such as the Observational Medical Outcomes Partnership Common Data Model are also popular.[35,36] As databases are brought together and made interoperable, researchers must advocate for improving the quality of health record data. EHR data can be of variable quality (eg, copying and pasting of text in clinical notes accounts for more than half of electronic notes).[37] The use of these flawed data to train AIs can perpetuate errors and biases.

Implementation details are vital as new approaches such as AI and ML are integrated into clinical immunology. Some practical lessons for integrating AI into the clinical workflow can be learned from sites that have had initial successes. The Vall d'Hebron Barcelona Hospital Campus is one such site, and as explained by Dr Xabier Michelena Vegas, an IT Healthcare consultant at Vall d'Hebron, the hospital's efforts started with a team approach that included information technology professionals, physicians, researchers, data security professionals, and human subject protection officers. He also echoed the general principal in EHR research of protecting patient interests by anonymization, pseudonymization, or codification of data. He proposed that EHR data remain traceable through various software pipelines, and although ideally, the raw data should not leave the institution's sights, an on-premises approach is no longer feasible; instead, institutions need to use virtual machines and cloud computing. He mentioned that institutions will need to recognize that their conservative rules for data protection may not be in the interests of the rare disease community, which in general has expressed an interest in more information sharing at the expense of privacy. Multiple levels of data protection and sharing may be useful to maximize the interests of patients with rare and nonrare conditions.

EQUITY AND FUTURE DIRECTIONS OF IA IN IEIs

Equity issues in the use of AI and ML in EHR research are important. Access to diagnostic approaches using AI will need to ensure that all patients, regardless of socioeconomic status, social group, age, or geographic localization, have an opportunity to benefit.[38] Newborn screening has resulted in universal access for early diagnosis of severe combined immunodeficiencies, regardless of socioeconomics. But still, there is much work to be done.

The access to health care for patients with IEIs is neither universal nor equitable around the globe. Leire Solis, health policy and advocacy senior manager from the International Patient Organization for Primary Immunodeficiencies, emphasized that it is crucial that individuals with IEIs take part in AI research initiatives and push for better access. Inequities are being addressed through certain initiatives, such as the collaborations among various research groups, scientific societies, and patient registries, as highlighted by Dr Elisa Hierro Cascajares and Jose Alfredo Mendez Barrera of the Autonomous Technological Institute of Mexico.[39] Using data from across registries of the major scientific societies will be key to mitigating the prevailing Western-centric bias in research by providing phenotype data for IEIs across the globe.

SUMMARY

The capabilities of AI offer the potential for earlier diagnoses of IEIs and an enhanced quality of life for individuals affected by IEIs. Harmonization across various databases around the world will be needed to provide sufficient data to train AI algorithms and will require our moving from proprietary EHR software to embracing Open Science principles instead. AI approaches will also benefit from collaborations between clinical experts, AI research groups (inside and outside the community of those studying IEIs), patient advocacy organizations, and scientific societies. Impediments to successful implementation of AI include operating in compliance with a rapidly changing legal landscape, enforcing data security and personal privacy, ensuring high-quality data for our algorithms, and encouraging proper ethical oversight to mitigate biases. Regardless, AI promises to transform the field of clinical immunology in the near future.

Acknowledgments

We are grateful to the Barcelona Super Computing center for hosting. We acknowledge the generous sponsorship from Grifols, Takeda, Pharming, CSL, The Jeffrey Modell Foundation, the Immune Deficiency Foundation, and Elevare; these sponsors did not influence the content of this article. We also acknowledge the contributions of the other participants of this conference: Josep Lluis Martí, Jordi Piera, Jordan Orange, Chris Runken, Eleonora Gambineri, Luiza Campos, Pablo González, Kirk Roberts, Lisa Bastarache, Xabier Michelena, Helen Leavis, Nicholas Rider, Antonio Condino-Neto, Elisa Hierro Cascajares, José Alfredo Méndez Barrera, Anna Sala, Alexandro Aisa, Nuria Castany, and Marta Juan.

Abbreviations used

AI	Artificial intelligence
CVID	Common variable immune deficiency
EHR	Electronic health record
HPO	Human Phenotype Ontology
ICD	International Classification of Diseases
IEI	Inborn error of immunity
IgRT	Immunoglobulin replacement therapy

LASID	Latin American Society for Immunodeficiencies
LLM	Large language model
ML	Machine learning
NLP	Natural language processing

REFERENCES

1. 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. [PubMed: 30419968]
2. Prince BT, Thielen BK, Williams KW, Kellner ES, Arnold DE, Cosme-Blanco W, et al. Geographic variability and pathogen-specific considerations in the diagnosis and management of chronic granulomatous disease. Pediatric Health Med Ther 2020;11:257–68. [PubMed: 32801991]
3. Lugo Reyes SO, Condino-Neto A, Stepensky P. Chapter 54 - Global perspectives on primary immune deficiency diseases. In: Sullivan KE, Stiehm ER, editors. Stiehm's immune deficiencies. 2nd ed. Cambridge, MA: Academic Press; 2020. pp. 1129–42.
4. Seth N, Tuano KS, Chinen J. Inborn errors of immunity: recent progress. J Allergy Clin Immunol 2021;148:1442–50. [PubMed: 34688776]
5. Basel D, McCarrier J. Ending a diagnostic odyssey: family education, counseling, and response to eventual diagnosis. Pediatr Clin of North Am 2017;64:265–72. [PubMed: 27894449]
6. Rosell AM, Pena LDM, Schoch K, Spillmann R, Sullivan J, Hooper SR, et al. Not the end of the odyssey: parental perceptions of whole exome sequencing (WES) in pediatric undiagnosed disorders. J Genet Couns 2016;25:1019–31. [PubMed: 26868367]
7. Carmichael N, Tsipis J, Windmueller G, Mandel L, Estrella E. "Is it going to hurt?": the impact of the diagnostic odyssey on children and their families. J Genet Couns 2015;24:325–35. [PubMed: 25277096]
8. Modell V, Gee B, Lewis DB, Orange JS, Roifman CM, Routes JM, et al. Global study of primary immunodeficiency diseases (PI)-diagnosis, treatment, and economic impact: an updated report from the Jeffrey Modell Foundation. Immunol Res 2011;51:61–70. [PubMed: 21935653]
9. Modell V, Orange JS, Quinn J, Modell F. Global report on primary immunodeficiencies: 2018 update from the Jeffrey Modell Centers Network on disease classification, regional trends, treatment modalities, and physician reported outcomes. Immunol Res 2018;66:367–80. [PubMed: 29744770]
10. Routes J, Costa-Carvalho BT, Grimbacher B, Paris K, Ochs HD, Filipovich A, et al. Health-related quality of life and health resource utilization in patients with primary immunodeficiency disease prior to and following 12 months of immunoglobulin G treatment. J Clin Immunol 2016;36:450–61. [PubMed: 27091140]
11. Modell F, Puente D, Modell V. From genotype to phenotype. Further studies measuring the impact of a physician education and public awareness campaign on early diagnosis and management of primary immunodeficiencies. Immunol Res 2009;44:132–49. [PubMed: 19140027]
12. Cunningham-Rundles C, Sidi P, Estrella L, Doucette J. Identifying undiagnosed primary immunodeficiency diseases in minority subjects by using computer sorting of diagnosis codes. J Allergy Clin Immunol 2004;113:747–55. [PubMed: 15100683]
13. Keller MD, Pandey R, Li D, Glessner J, Tian L, Henrickson SE, et al. Mutation in IRF2BP2 is responsible for a familial form of common variable immunodeficiency disorder. J Allergy Clin Immunol 2016;138:544–50.e4. [PubMed: 27016798]
14. Orange JS, Glessner JT, Resnick E, Sullivan KE, Lucas M, Ferry B, et al. Genome-wide association identifies diverse causes of common variable immunodeficiency. J Allergy Clin Immunol 2011;127:1360–7.e6. [PubMed: 21497890]
15. Bastarache L Using phecodes for research with the electronic health record: from PheWAS to PheRS. Annu Rev Biomed Data Scie 2021;4:1–19.

16. Bastarache L, Hughey JJ, Hebbring S, Marlo J, Zhao W, Ho WT, et al. Phenotype risk scores identify patients with unrecognized mendelian disease patterns. Science 2018;359:1233–9. [PubMed: 29590070]

17. Cassini T, Bastarache L, Zeng C, Han ST, Wang J, He J, et al. A test of automated use of electronic health records to aid in diagnosis of genetic disease. Genet Med 2023;25:100966. [PubMed: 37622442]

18. Shuey M, Stead W, Aka I, Barnado A, Bastarache J, Brokamp E, et al. Next-generation phenotyping: introducing PhecodeX for enhanced discovery research in medical phenomics. Bioinformatics 2023;39:btad655. [PubMed: 37930895]

19. Aref L, Bastarache L, Hughey JJ. The phers R package: using phenotype risk scores based on electronic health records to study mendelian disease and rare genetic variants. Bioinformatics 2022;38:4972–4. [PubMed: 36083022]

20. Bastarache L, Delozier S, Pandit A, He J, Lewis A, Annis AC, et al. The phenotype-genotype reference map: improving biobank data science through replication. Am J Hum Genet 2023;110:1522–33. [PubMed: 37607538]

21. Rivière JG. The PIDCAP project: developing a warning-sign-based algorithm for use in an electronic health record screening tool for inborn errors of immunity screening. Rotterdam, The Netherlands: International Primary Immunodeficiencies Congress; Nov 8-10, 2023.

22. 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. [PubMed: 30937298]

23. 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. [PubMed: 33591972]

24. Rider NL, Cahill G, Motazedi T, Wei L, Kurian A, Noroski LM, et al. A risk prediction and clinical guidance system for evaluating patients with recurrent infections. medRxiv 2020. 10.1101/2020.06.12.20129692.

25. 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. [PubMed: 36243223]

26. Johnson R, Stephens AV, Knyazev S, Kohn LA, Freund MK, Bondhus L, et al. Electronic health record signatures identify undiagnosed patients with common variable immunodeficiency disease. Health Informatics. 2022. Available at: http://medrxiv.org/lookup/doi/10.1101/2022.08.03.22278352. Accessed November 27, 2023.

27. Fang H, Frean M, Sylwestrzak G, Ukert B. Trends in disenrollment and reenrollment within US commercial health insurance plans, 2006–2018. JAMA Netw Open 2022;5:e220320. [PubMed: 35201308]

28. Schiavo E, Martini B, Attardi E, Consonni F, Ciullini Mannurita S, Coniglio ML, et al. Autoimmune cytopenias and dysregulated immunophenotype act as warning signs of inborn errors of immunity: results from a prospective study. Front Immunol 2021;12:790455. [PubMed: 35058929]

29. Rider N, Shin H, Roberts K. Text-mined data improves machine learning predictions for detecting inborn errors of immunity. Clinical Immunol 2023;250:109602.

30. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 2015;36: 928–30. [PubMed: 26220891]

31. Zurek B, Ellwanger K, Vissers LELM, Schüle R, Synofzik M, Töpf A, et al. Solve-RD: systematic pan-European data sharing and collaborative analysis to solve rare diseases. Eur J Hum Genet 2021;29:1325–31. [PubMed: 34075208]

32. Bernstam EV, Warner JL, Krauss JC, Ambinder E, Rubinstein WS, Komatsoulis G, et al. Quantitating and assessing interoperability between electronic health records. J Am Med Inform Assoc 2022;29:753–60. [PubMed: 35015861]

33. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the crisis into an opportunity: digital health strategies deployed during the COVID-19 outbreak. JMIR Public Health Surveill 2020;6:e19106. [PubMed: 32339998]

34. Cunha J, Duarte R, Guimarães T, Santos MF. OpenEHR and business intelligence in healthcare: an overview. Procedia Computer Science 2023;220:874–9.

35. Park K, Cho M, Song M, Yoo S, Baek H, Kim S, et al. Exploring the potential of OMOP common data model for process mining in healthcare. PLoS One 2023;18: e0279641. [PubMed: 36595527]

36. Biedermann P, Ong R, Davydov A, Orlova A, Solovyev P, Sun H, et al. Standardizing registry data to the OMOP common data model: experience from three pulmonary hypertension databases. BMC Med Res Methodol 2021;21:238. [PubMed: 34727871]

37. Steinkamp J, Kantrowitz JJ, Airan-Javia S. Prevalence and sources of duplicate information in the electronic medical record. JAMA Netw Open 2022;5: e2233348. [PubMed: 36156143]

38. Lawrence MG, Rider NL, Cunningham-Rundles C, Poli MC. Disparities in diagnosis, access to specialist care, and treatment for inborn errors of immunity. J Allergy Clin Immunol Pract 2023;S2213219823011960.

39. Méndez Barrera JA, Rocha Guzmán S, Hierro Cascajares E, Garabedian EK, Fuleihan RL, Sullivan KE, et al. Who's your data? Primary immune deficiency differential diagnosis prediction via machine learning and data mining of the USIDNET registry. Clin Immunol 2023;255:109759. [PubMed: 37678719]

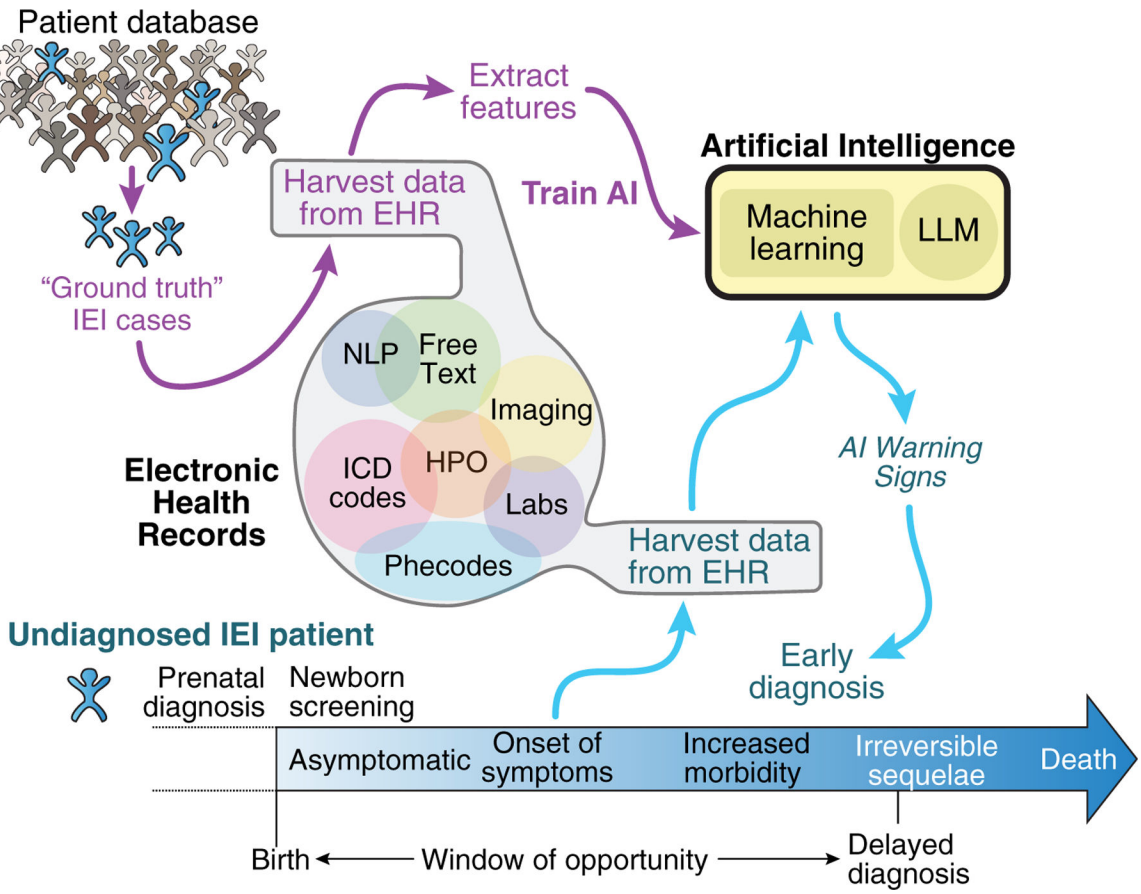

FIG 1.
Opportunity for AI to accelerate the diagnosis of IEIs.