Patient-level AI Collaboration for Precision Medicine in Canada: A Scoping Review

Omid Jafarinezhad

o.jafarinezhad@mail.utoronto.ca

Erica Zhang
Ryan Rezai
Mohammad Noaeen
Aviv Shachak
Behrouz Far
Zahra Shakeri

zahra.shakeri@utoronto.ca

Systematic Review

Keywords:

Posted Date: September 16th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7622437/v1

Additional Declarations: The authors declare no competing interests.

Patient-level AI Collaboration for Precision Medicine in Canada: A Scoping Review

Omid Jafarinezhad[1], Qian Zhang[1], Ryan Rezai[1,2], Mohammad Noaeen[1], Aviv Shachak[1,3], Behrouz Far[4], and Zahra Shakeri[1,3,5,*]

[1]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
[2]Department of Electrical and Computer Engineering, McGill University, Montréal, Canada
[3]Faculty of Information, University of Toronto, Toronto, Canada
[4]Department of Electrical and Software Engineering, University of Calgary, Calgary, Canada
[5]Schwartz Reisman Institute, University of Toronto, Toronto, Canada
[*]Corresponding author, zahra.shakeri@utoronto.ca

Precision medicine runs on patient-level data. In Canada, most patient-level AI collaborations are international, and many draw on datasets housed outside the country. We mapped the landscape of Canadian healthcare-AI collaborations using patient-level data and found that privacy-preserving data-sharing practices are extremely rare. Using a custom large language model pipeline, we screened 245,886 articles (2018–Feb 2025) and identified 3,100 relevant studies. The main dataset drawn from PubMed after a pilot assessment against IEEE Xplore, ACM Digital Library, Scopus, and Web of Science indicated equivalent coverage for our eligibility criteria. We identified trends in domestic and international collaborations and examined implications for data sharing, privacy, and interoperability. Among 3,100 studies, only 35.8% were conducted solely by Canadian institutions and nearly two thirds involved international partners. Imaging data were the most common modality. Alarmingly, out of 160 multi-site collaborations, just 5% (8 studies) used any privacy-preserving method, and only one of those was a fully within-Canada collaboration using Canadian patients' data. Canada's slow adoption of privacy-preserving collaboration technologies is limiting the country's ability to fully leverage AI in healthcare. Without immediate investment in secure, decentralized data sharing infrastructure, Canada will fall behind in health AI. Prioritizing such infrastructure is critical to enable innovation and to ensure Canadian patient data benefit Canadian patients.

Introduction

Patient-level data allow models to capture clinical complexities, support local validation, and inform care pathways. When access depends on foreign repositories or ad hoc exchanges, Canadian studies may not reflect Canadian populations, which limits clinical utility. A national view of how collaborations are organized is needed to guide responsible and patient-centered AI in Canada. Collaboration enables investigators to integrate complementary expertise, data, and infrastructure.[1–3] It expands interdisciplinary capacity to tackle questions that isolated teams cannot address.[3,4] Over the past decade, high-income countries in the Organization for Economic Co-operation and Development (OECD) have invested in multicenter consortia that share data and analytic capacity, especially for precision-health initiatives.[5] Canada endorses this agenda in policy statements, however national reviews indicate that laboratories exchange patient-level datasets less frequently than international counterparts.[5,6] Geographic variation, provincial stewardship of health records, variable privacy legislation, and uneven digital infrastructure create barriers to joint work.

Modern medical research relies on data sharing, the transfer of data between investigators at different sites. Top journals now require it to get more value from costly datasets and to speed discovery.[7] In healthcare, however, moving data across sites is unusually difficult. Clinical records are sensitive and subject to strict legal, ethical, and technical obligations that limit cross-institution transfer. These safeguards protect participants and reduce harm, and the constraints tighten as data become more granular, with patient-level data carrying the highest risk and the most stringent requirements.[8] *Patient-level data* are observations on individual participants that carry a risk of re-identification.[9] Such data are used to train most medical artificial intelligence (AI) models, including electronic health records (EHRs), radiological images (MRI, CT, X-ray), biosignals (ECG, EEG, EMG), and genomic sequences.[10] Models that train on these large and detailed datasets can detect patterns unreachable by conventional statistics, which supports precision care at the scale of a single person.[11] Patient-level resources have therefore given a boost to the rapid spread of AI across virtually every discipline of medicine.[12] However, limited access to heterogeneous Canadian datasets jeopardizes equity, as algorithms developed on narrow or homogeneous samples may embed bias and perform inconsistently across the country's diverse population.

The need for patient-level data is increasing with the rise of medical foundation models, which are trained on large datasets to perform multiple clinical tasks.[13] Some go further, some medical foundation models are trained on more than one modality of patient-level data.[14–16] These models show potential across many clinical tasks, but their development depends on collaboration between universities, hospitals, and research centers that can supply complementary datasets. Canada's fragmented data governance constrains such collective efforts and slows progress toward equitable AI. Policymakers have responded to these barriers through Bill C-72, introduced in 2024, which would establish a framework for secure, interoperable

medical data exchange and prohibits data blocking by health information vendors.[17] The success of this legislation will depend on a clear picture of existing collaboration patterns and practical technical options that can deliver privacy-preserving sharing at scale.

Despite recognition of the issue, there is no systematic account of how Canadian researchers share patient-level data or how often and with whom they partner internationally. We conducted a systematic scoping review of Canadian practices from 2018 to February 2025, using a mixed-methods approach that combined large-scale bibliometric analysis with a qualitative review of technical and policy contexts.[18] Our quantitative analyses map collaboration ties, where they form, and which data types are most common (e.g., MRI and CT). Our qualitative analyses contextualize these patterns by examining the ethical, regulatory, and sociotechnical factors that shape collaboration. Together, the two strands offer an integrated view of collaboration and its technical, ethical, and policy drivers.

Guided by sociotechnical systems theory, which looks at how technology and organization fit together,[19,20] we examine how collaborative networks combine AI tools with privacy-preserving methods to advance patient-level research while protecting confidentiality.[21] From these findings, we outline practical steps to broaden equitable access to diverse datasets and to strengthen Canada's contributions to precision medicine. We also flag the priorities for the next steps, including underused data sources, key standards, and coordination mechanisms.

Results

We executed a comprehensive, iterative search that combined database queries with manual screening to compile Canadian studies on collaborative medical AI. The review protocol, selection process, and data extraction followed the Preferred Reporting Items for Systematic Reviews and Meta analyses Extension for Scoping Reviews (PRISMA-ScR) checklist,[22] which provided an appropriate reporting framework for our objectives compared with alternative approaches.[23–25] The PubMed search retrieved 245,886 records, restricting the set to publications with at least one Canadian-affiliated author yielded 9,238 items. After removing 310 editorials/letters and 190 irrelevant reviews, 8,738 records remained for title-abstract screening. We excluded 4,790 studies that lacked patient-level data, leaving 3,948 articles for assessment. A further 848 articles were removed due to irrelevant topics (769 articles) or use of only public/synthesized datasets (79 articles), resulting in 3,100 studies for qualitative synthesis (Figure 1). Within this final corpus, 160 articles described multicenter data collection efforts: 83 incorporated Canadian data alongside other sources, 32 relied exclusively on Canadian data, and 45 did not contain any Canadian data.

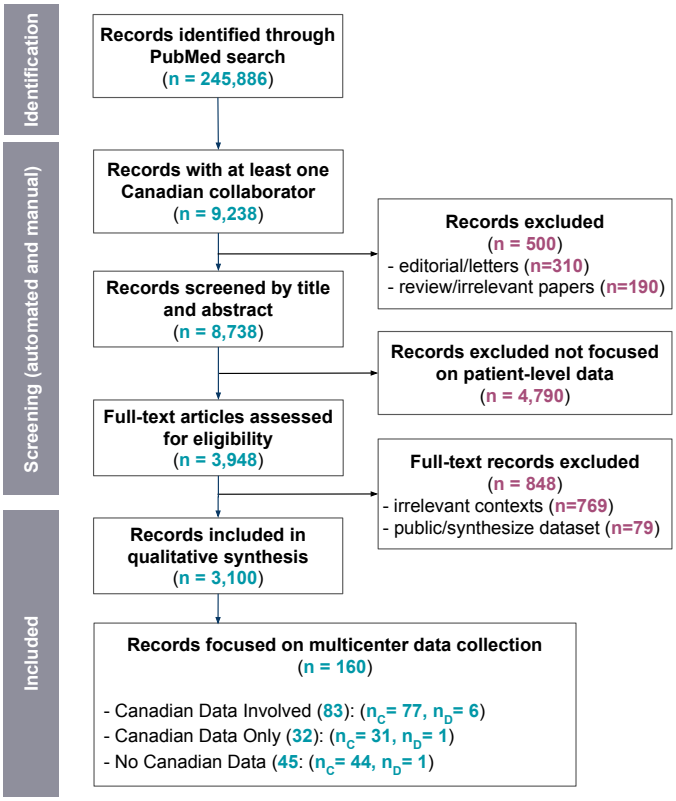
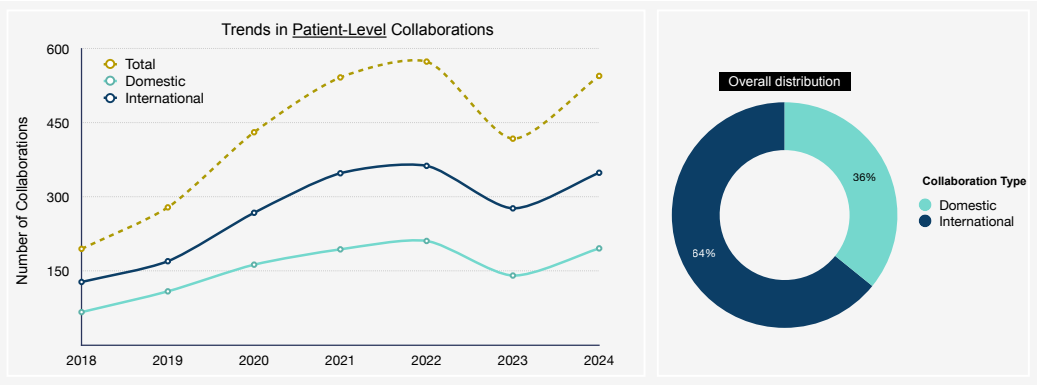


Figure 1: The overall process of article selection following PRISMA-ScR guideline. The final box highlights the breakdown of Canadian data involvement: among 115 studies with Canadian patient data, 83 included Canadian data alongside other countries (77 using centralized repositories (n_C), 6 using decentralized approaches (n_D)), and 32 studies used exclusively Canadian data (31 centralized, 1 decentralized). Among the 45 studies without Canadian data, 44 used centralized approaches and 1 used a decentralized approach.

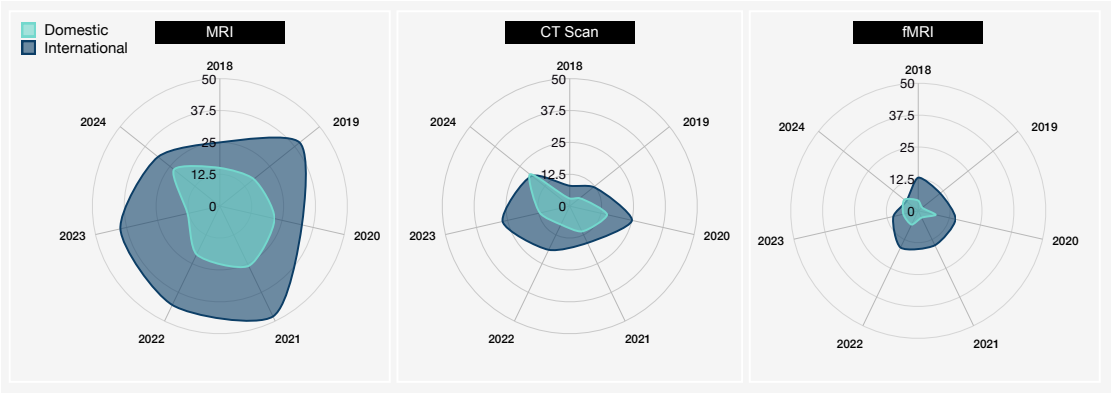
Trends in patient-level Data Collaboration

We first examined collaboration trends by institution and country, categorizing each study as domestic or international. Domestic collaborations refer to studies where all authors are affiliated with Canadian institutions, whereas international collaborations involve at least one Canadian-affiliated author in partnership with an institution outside Canada. For example, a Canada-only study on oral cancer surgery outcomes was conducted across two Toronto hospitals (i.e., University Health Network (UHN) and Sunnybrook),[a1] whereas an international consortium analyzed white-matter abnormalities in obsessive–compulsive disorder (OCD) from 18 sites across 16 countries as part of the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) initiative.[a2]

We found a strong trend toward international collaborations, with nearly two-thirds of Canadian-linked projects were partnered with foreign institutions. Canadian teams frequently relied on US data sources and global networks, showing a dependence on external partners. For example, a study of amyotrophic lateral sclerosis MRI combined data from Canada and the United States (e.g., Miami and Utah)[a3] , and an international deep reinforcement learning study optimized warfarin dosing in atrial fibrillation using data from 52 countries, with external validation in 44 countries,[a4] and these cases make the dependence clear. It is also consistent with observations of the Pan-Canadian Health Data Strategy Expert Advisory Group, which has documented systemic fragmentation and persistent interjurisdictional barriers in Canada's health data landscape.[26]



(a) Annual counts of domestic and international (Canada-non-Canada) collaborations using patient-level data with overall proportions.



(b) Domestic and international collaborations by imaging modality (MRI, CT, fMRI).

Figure 2: **Collaborations involving Canadian institutions that use patient-level data.** Domestic collaborations are projects conducted entirely between Canadian partners, whereas international collaborations include at least one partner outside Canada: (a) shows yearly collaboration counts from 2018 to 2024 together with their overall proportions, and (b) presents the distribution of collaborations across three imaging modalities (MRI, CT, and fMRI) for both domestic and international categories. Note that visualizations are presented for full calendar years (2018–2024) to ensure consistency and comparability.

Of the 3,100 collaborations in our dataset, 1,989 (64.2%) involved international partners, while 1,111 (35.8%) were conducted solely among Canadian institutions (Figures 2a and 2b).

Annual collaboration counts rose in parallel for both domestic and international studies between 2018 and 2022, suggesting that COVID-19 targeted funding initiatives and the rapid adoption of secure cloud collaboration tools stimulated international and domestic projects equally. The total of collaborations increased steadily from 195 in 2018, to a peak of 574 in 2022. The activity then decreased to 418 in 2023 and rebounded to 546 in 2024 (Figure 2a), likely reflecting a temporary reallocation of research efforts towards pandemic recovery in 2023 and the subsequent renewal of federal AI funding in 2024. A sensitivity analysis restricted to studies that passed the assessment (n = 3,100) reproduced this trajectory, supporting the robustness of our automated LLM-based screening pipeline. The inflection after 2022 corresponds to reports of a broader post-pandemic contraction in scientific output, due to reduced research time, shifts in funding priorities, and resource constraints,[27] as well as the resumption of routine clinical services that diverted personnel from research activities.

International collaborations consistently made up the majority of patient-level data studies throughout the period, peaking at 363 in 2022 before decreasing to 349 in 2024. Domestic collaborations followed a similar pattern, rising to 211 in 2022 and then dropping to 197 by 2024, despite targeted national programs (e.g., the Canadian Institutes of Health Research (CIHR) Catalyst Grants) intended to stimulate intra-provincial data consortia. This downturn in 2023 coincided with the end of that program's pilot phase and the redeployment of several provincial analytics teams to clear pandemic-related backlogs. Because this scoping review excluded studies that used only public or synthetic data, the collaborations we captured primarily include projects that required access to patient records governed by external institutions. In practice, these collaborations face substantial governance challenges such as data localization requirements and cross-border consent processes. Our qualitative analysis of the full text methods sections revealed that 81.8% of the multicenter projects required parallel ethics board approvals in at least three different jurisdictions, greatly increasing administrative costs and timelines. For example, a 13 site ENIGMA Bipolar consortium study coordinated local approvals in each institution before combining structural MRI data for machine learning,[a5] and a German and Canadian kidney transplant project obtained separate REB clearances in Montréal, Vancouver, and Berlin.[a6]

In the multicenter subset, only 32 of the 160 collaborations (20%) relied exclusively on data from Canadian patients,

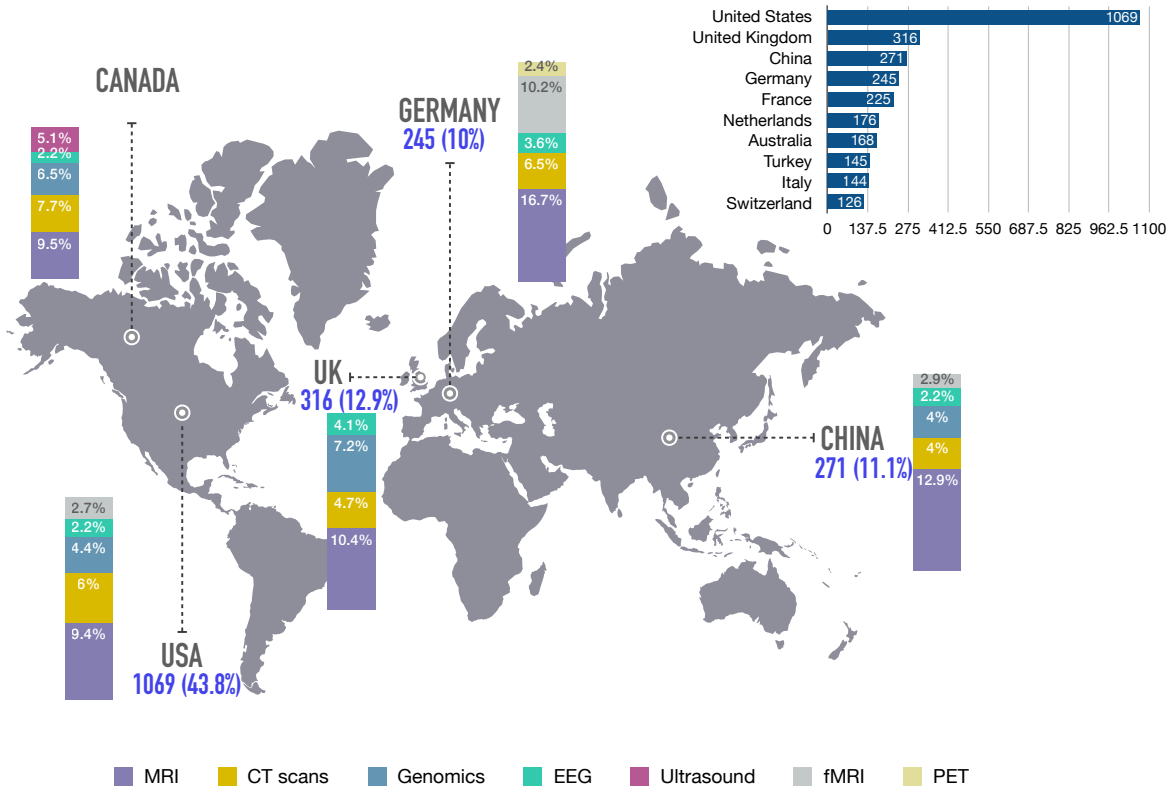


Figure 3: **International Collaborations in Medical AI Involving Canada from 2018 to 2024 (full calendar years)**. Geographic distribution of patient-level data collaborations between Canada and international partners. The bar chart (top-right) ranks the ten most frequent partner countries, while stacked bars beside selected nations show the five most common data modalities exchanged with Canada (e.g., Canada-all partners, Canada-US).

while 128 (80%) included data from foreign sites. More than 40% of the international partnerships in this subset involved data from the United States (43.8%), making it the most frequent non-Canadian data source. This pattern aligns with policy frameworks that encourage interoperable, cross-border data exchange, including the USMCA's digital-trade provisions and the G7 Health Ministers' 2021 commitments on open standards and the International Patient Summary.[28,29] Without a national trusted-access framework, Canadian researchers often must depend on foreign repositories to obtain the diversity of data their models require. A 2025 OECD review noted that fragmented approval processes in a Canadian hospital stretched a cross-border data-access effort to 18 months, until process re-engineering shortened it to less than 3 months.[30] Such cases highlight the importance of translating these lessons into a cohesive Canadian governance roadmap that provides timely and domestic access to richly varied patient data.

Data Modalities

Figure 3 provides a geographic overview of Canada's international collaborations, highlighting the most frequent partner countries and the types of data exchanged. Across collaborations from 2018-2024, imaging was the most common primary analytical modality, while clinical variables were used almost universally but mainly as supporting covariates. Studies often shared radiological images such as MRI and CT scans, including functional MRI (fMRI), which appeared in more than 22% of all collaborations. For example, national brain tumor consortia pooled MRI scans from dozens of sites,[a7] and a Canada-US lung cancer screening study shared low-dose chest CT images to develop radiomics models.[a8] The geographic breakdown of partnerships by modality further shows that major international partners (e.g., the United States with 43.8% of partnerships, the United Kingdom 12.9%, China 11.1%, Germany 10.0%) often exchanged imaging datasets (Figure 3). MRI was the most common data modality in our dataset. The prominence of medical imaging likely reflects the rich diagnostic information these data provide, the long-standing use of standardized formats such as DICOM that simplify integration between institutions, and the compatibility of images with advanced AI techniques. The popularity of imaging data could also relate to an emphasis on model interpretability in clinical AI. For example, methods such as Grad-CAM enable visualization of image-based model predictions,[31] which help build clinicians' trust in the AI's outputs.

Although medical images are widely used, they pose significant challenges for AI research. Annotating imaging data at high quality is labor intensive and time consuming, and strict privacy regulations can limit the sharing of sensitive scans, making it difficult to build large multi-institutional datasets.[32] In addition, many hospital picture archiving and communication systems store images in proprietary formats that require substantial preprocessing and curation before they can be used for model development. Other data modalities, such as genomics and biosignals (e.g., EEG, ECG), appeared less fre-

quently in our review but hold important promise for precision medicine. Genomic data can enable personalized diagnostics and targeted therapies, while biosignals support real-time patient monitoring. However, these modalities bring their own challenges. Genomic datasets are typically high-dimensional and require complex preprocessing and interpretation, making them harder for AI models to handle effectively. Similarly, biosignals are often noisy and variable, which complicates the extraction of consistent features and insights.[33] As multimodal AI approaches mature, we anticipate that more studies will incorporate these underused modalities once robust processing and integration pipelines are established.

Multicenter Data Collection Collaborations

As noted earlier, among the 160 multicenter data collection collaborations, 32 (20%) were domestic (Canada-only) and 128 (80%) included at least one international site. The annual trend for these multicenter projects is shown in Figure 4, divided into domestic and international categories. International multicenter collaborations began to rise sharply after 2020, reaching 43 studies in 2024. This increase may reflect a growing global emphasis on cross-border research partnerships in healthcare, potentially accelerated by the COVID-19 pandemic, which highlighted the need for international collaboration to address global health challenges. Domestic multicenter collaborations also increased during the study period, but at a slower pace. These Canada-only projects peaked in 2023 and then declined slightly. Overall, the total volume of multicenter collaborations (i.e., domestic plus international) rose sharply starting in 2022, driven mainly by the growth of international partnerships. Detailed domain-, modality-, and task-level summaries of these multicenter collaborations are provided in Tables 1 and 2.

Consistent with the broader dataset, in which imaging appeared in over 22% of studies, more than half of multicenter collaborations incorporated imaging data (82, 51.2%), including MRI in 30 (18.7%) and CT in 25 (15.6%). In most cases, imaging was used alongside other data rather than alone (31, 19.4% image-only vs 53, 33.1% multimodal), showing its role as a key analytical modality frequently paired with clinical and laboratory information. Clinical data appeared in 102 (63.8%), but only 19 (11.9%) studies were clinical-only and 83 (51.9%) were multimodal. Genomic data appeared in 14 (8.7%) studies. This distribution reinforces the dominance of imaging techniques in collaborative research, while also showing an emerging, though still limited, role for genomic data in multi-site studies, consistent with the broader shift toward precision medicine and genomics-driven care. For example, a multinational multicenter lung-transplant study involving 10 centers in six countries used microarray-based gene expression profiling and machine learning to derive biopsy-based molecular T-cell-mediated rejection (TCMR) scores, with higher scores associated with subsequent loss of graft.[a9] Likewise, a pharma-academic alliance assembled a centralized, research ready database that integrates clinical records, MRI, and omics across large multicenter trial cohorts (e.g., 35,000 multiple sclerosis (MS) patients and >230,000 MRIs), allowing imaging omics analyses at scale.[a10]

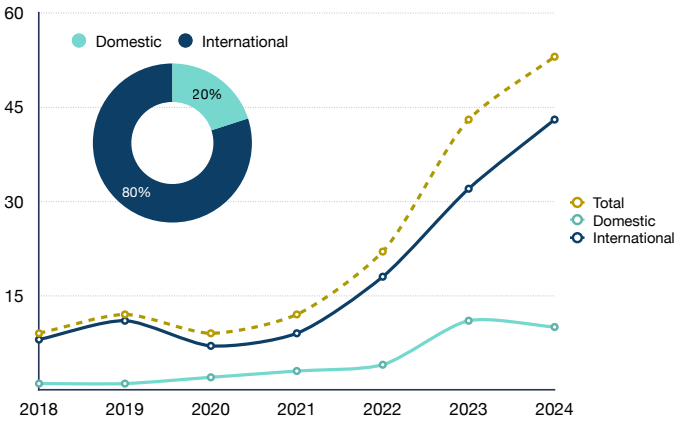


Figure 4: **Trends in Domestic and International Multicenter Collaborations, from 2018 to 2024 (full calendar years).** Annual counts of Canadian-led multicenter collaborations are shown for domestic, international, and combined totals. The inset doughnut summarizes the full period: 80% international and 20% domestic. Overall activity rises sharply after 2022, driven mainly by international partnerships, while domestic collaborations peak in 2023 before declining slightly.

We profiled data modalities and disease relationships with a co-occurrence analysis and hierarchical clustering. First, we standardized terminology across studies to group related data types. For example, terms such as 'genomic sequencing' and 'gene expression' were combined under the broader label 'Genomic Data'. Similarly, patient-reported outcomes and quality-of-life measures were grouped as 'Survey Data'. We then constructed a binary matrix of the presence of modality in disease categories and applied hierarchical clustering to the resulting co-occurrence matrix to reveal patterns of modality use (Figure 5).

Clustering the modalities alongside disease categories revealed several patterns. Cancer & Oncology (56) spanned multiple modalities but relied most on imaging (25), followed by clinical (18) and laboratory data (10). Cardiovascular and Vascular Diseases (45) combined clinical (19) and imaging (17) data, with distinctive use of biosignals (4). Psychiatric and Mental Health (32) relied heavily on survey data (8), alongside imaging (10) and clinical (9), and was one of only two areas that used environmental data. Neurological Disorders (30) showed the broadest mix, with activity in seven modalities. Although used less frequently, genomic and molecular data appeared in certain specialized areas, providing important insights in Surgical and Trauma Conditions as well as Psychiatric and Mental Health collaborations where genetic information supports personalized treatment. Laboratory biomarkers were also central in Cancer and Oncology and Kidney and Renal Diseases, where test results guide diagnosis and monitoring. These findings indicate a widespread reliance on imaging and electronic health record data across disease domains, together with an increasing use of genomic and laboratory data in areas that benefit from molecular perspectives.

Table 1: Summary of multicenter AI collaborations with Canadian participation (2018–2024).

Domain	Modality mix	Collab scope-data mgmt(#)	Privacy tech	Explainability	Primary tasks	Interop / deploy
Cancer & Oncology[a1,a7,a8,a11–a39]	Clinical Records[a1,a13,a14,a17 a8,a20,a21,a24–a30,a33,a34,a36,a38] Imaging Data[a7,a11,a12,a14–a21,a23 a8,a25–a32,a35,a37–a39] Lab Data[a17,a22,a25,a30,a32,a33 a11,a34–a36] Survey Data[a13,a25] Genomic Data[a11]	CA–C(2),[a1,a13,a31] NoCA–C(pooled),[a11] Intl–D(6),[a12] Intl–D(71),[a7] NoCA–C(11),[a14] Intl–C(3),[a15,a21] NoCA–C(3),[a16,a22] Intl–C(15),[a17] Intl–C(5),[a18–a20,a32,a36] NoCA–C(2),[a23] Intl–C(pooled),[a24] Intl–C(6),[a25] Intl–C(2),[a26,a28,a30] Intl–C(4),[a8,a27] CA–C(4),[a29,a37] NoCA–CA(6),[a33] NoCA–C(5),[a34] CA–C(5),[a35] Intl–C(12),[a38] NoCA–C(3)[a39]	Privacy-preserving distributed learning[a12] Federated learning[a7] REDCap for secure data capture[a13]	Global interpretability[Gini importance;[a1] transparent coefficients;[a8,a21,a22,a33,a33,a34] RF importance,[a14,a15,a26,a30] Feature ranking[a31]], SHAP explanations[a13,a38], Imaging saliency [CAM,[a19] Heatmap localization (AIDA)[a37]]	Clf[a11,a12,a14,a15,a19,a20,a22,a30 a8,a32,a35,a37–a39] Pred[a1,a12,a16,a17,a21,a23,a26–a29,a3 a8,a33,a34,a36] Seg[a7,a18] Risk/Stratify[a13,a17,a21,a24,a32,a33] Enablement/Infra[a25]	Shared code (GitHub) [a7,a8,a26] Web tools [a21]
Cardiovascular & Vascular Diseases[a4,a40–a64]	Clinical Records [a40,a42,a45 a4,a47–a52,a55,a57–a64] Imaging Data[a41–a44,a46–a49 a51,a52,a54–a56,a58,a59,a62,a63] Lab Data[a45,a50,a54,a61] Biosignal Data[a53,a61,a62,a64] Genomic Data[a60]	NoCA–C(+700),[a40] NoCA–C(4),[a41] Intl–C(38),[a42] Intl–C(9),[a43,a48] Intl–C(2),[a44,a62] CA–C(3),[a45] Intl–C(12),[a46] Intl–C(3),[a47,a58] CA–C(2),[a49,a59] NoCA–C(2),[a50,a54] Intl–C(5),[a51,a52] NoCA–C(15),[a53] Intl–C(29),[a55] Intl–C(10),[a56] CA–C(9),[a57] Intl–C(44),[a4] CA–C(12),[a60] Intl–C(4),[a61,a63] CA–D(7)[a64]	Federated learning + differential privacy[a64]	SHAP explanations [a54,a57], Feature importance [XGBoost gain,[a40,a45] information gain,[a48] RF/GBM importance[a50]], Imaging saliency [Saliency mapping of ECG traces[a62]]	Clf[a41,a43,a45,a46,a48,a53,a64] Pred[a4,a40,a42,a50,a56–a62] Seg[a44] Risk/Stratify[a47,a49,a51,a52 a54,a55,a57,a63]	Shared code (GitHub) [a40,a60,a64] Web tools [a4]
Neurological Disorders[a3,a65–a83]	Clinical Records [a67,a71,a73 a75,a79,a80,a83] Imaging Data[a66,a68–a72,a74 a3,a76–a78,a81–a83] Lab Data[a79] Digital Devices[a65,a75] Biosignal[a66,a68,a70] Genomic Data[a80] Survey Data[a67,a82]	NoCA–C(4),[a65,a76] Intl–C(25),[a66] Intl–C(12),[a67] Intl–C(3),[a68] Intl–C(30),[a69] Intl–C(608),[a70] NoCA–C(3),[a71] NoCA–C(pooled),[a72] Intl–C(pooled),[a73] NoCA–C(2),[a74,a75] Intl–C(5),[a3] Intl–C(2),[a77,a82] CA–C(24),[a78] CA–C(31),[a79] Intl–C(4+),[a80] Intl–D(83),[a81] Intl–C(22)[a83]	Client-side encryption + server-side encrypted storage[a65] Distributed learning[a81]	Global interpretability [a65,a67,a71,a74,a75,a83] SHAP explanations [a70,a79] Imaging saliency [Saliency maps (SmoothGrad)[a72]]	Clf[a3,a65,a68,a71,a78,a81] Pred[a65,a67,a70,a72,a75,a77,a82,a83] Seg[a69,a76] Risk/Stratify[a66,a73–a75,a77,a79 a80,a82]	Shared code (GitHub) [a77,a81,a82]
Psychiatric & Mental Health[a2,a5,a84–a99]	Clinical Records [a88–a90,a92–a96,a99] Imaging Data [a2,a5,a86–a88,a91–a95] Survey Data [a84–a86,a90,a95 a97–a99] Genomic Data [a84,a89,a94] Lab Data [a87] Environment Data [a84]	NoCA–C(18),[a2] Intl–C(13),[a5] NoCA–C(21),[a84] Intl–C(2),[a85] Intl–C(8),[a86] Intl–C(36),[a87] CA–C(6),[a88] NoCA–C(10),[a89] NoCA–C(6),[a90] Intl–C(30),[a91] Intl–C(7),[a92] Intl–C(21),[a93] NoCA–C(3),[a94] Intl–C(3),[a95,a99] CA–C(17),[a96] Intl–C(29),[a97] Intl–C(40)[a98]		Global interpretability [a2,a5,a84–a87,a89–a91,a93,a94,a96,a97,a99] Local interpretability [SHAP,[a92] K-LIME[a2]]	Clf[a2,a5,a87,a91] Pred[a85,a86,a88–a90,a93,a94,a96,a99] Risk/Stratify[a84,a92,a95,a97,a98]	Shared code (GitHub) [a84,a96] ENIGMA's harmonized processing [a2,a5,a87]
Respiratory & Lung Conditions[a100–a112]	Clinical Records [a100,a101 a104,a105,a107,a109,a112] Imaging Data [a102,a103,a106 a108,a111,a112] Survey Data [a104] Genomic Data [a100,a106] Lab Data [a100,a107] Biosignal Data [a110] Environment Data [a104]	NoCA–C(8),[a100] Ca–C(48),[a101] CA–C(2),[a102,a107] NoCA–C(2),[a103,a112] Int-C(4),[a104] NoCA–C(105),[a105] NoCA–C(4),[a106] Intl–C(3),[a108] Intl–C(785),[a109] CA–C(3),[a110] Intl–C(5)[a111]		Global interpretability [XGB gain;[a101] LR coeffs, tree MDI, SVM permutation[a104]] Imaging saliency [Grad-CAM;[a102] CAM;[a103] Grad-CAM++[a108]]	Clf[a101–a104,a108,a110,a111] Pred[a107] Risk/Stratify[a100,a105,a106,a109,a11	Shared code (GitHub) [a111]
Kidney & Renal Diseases[a6,a113–a118]	Clinical Records [a6,a113–a118] Genomic Data [a6] Lab Data [a6,a113–a118]	Intl–C(4),[a6] Intl–D(4),[a116] Intl–C(3),[a113] CA–C(2),[a114] CA–C(201),[a115] CA–C(pooled),[a117] Intl–C(24)[a118]	Privacy-preserving FL and private blockchain (Ethereum DLT)[a6]	Global interpretability [a114]	Clf[a6] Pred[a6,a113,a114,a117] Risk/Stratify[a115,a116,a118]	Docker packaging and REST API[a6]
Infectious Diseases[a119–a124]	Clinical Records [a119 a121–a124] Imaging Data [a119,a120] Lab Data [a121–a124] Genomic Data [a122]	Intl–D(20),[a119] NoCA–C(5),[a120] CA–C(7),[a121] NoCA–C(4),[a122] CA–C(19),[a123] Intl–C(350+)[a124]	FL (FedAvg)[a119]	Imaging saliency [CAM[a120]]	Pred[a119,a120,a122] Risk/Stratify[a121,a123] Enablement/Infra[a124]	Shared code (GitHub) [a120] Released on NVIDIA NGC (Clara Train)[a119]

Abbreviations: **CA** Canada-only; **Intl** international (including Canada); **NoCA** No Canadian data; **C** centralized; **D** decentralized; **FL** federated learning; **Clf** Classification; **Pred** Prediction; **Seg** Segmentation.

Table 1 (continued): Summary of multicenter AI collaborations with Canadian participation (2018–2024).

Domain	Modality mix	Collab scope-data mgmt(#)	Privacy tech	Explainability	Primary tasks	Interop / deploy
Critical Care[a125–a129]	Clinical Records [a125–a129] Digital Devices [a128,a129] Lab Data [a125–a129]	NoCA–C(pooled),[a125] Intl–C(20),[a126] CA–C(pooled),[a127] NoCA–C(4),[a128] CA–C(16)[a129]		Global interpretability [a126,a128]	Clf [a125,a127,a129] Pred [a126,a128,a129]	Shared code (GitHub) [a129]
Surgical & Trauma Conditions[a9,a130–a133]	Clinical Records [a131–a133] Imaging Data [a131] Genomic Data [a9,a130,a132]	Intl–C(10),[a9] Intl–C(7),[a130] NoCA–C(20),[a131] Intl–C(8),[a132] NoCA–C(3)[a133]		Global interpretability [a9,a131], SHAP explanations [a133]		Web page calculator[a133]
Reproductive, Pregnancy, and Neonatal Health[a134–a138]	Clinical Records [a134–a138] Lab Data [a135,a136,a138]	NoCA–C(10),[a134] Intl–C(3),[a135] Intl–C(17),[a136] Intl–C(49),[a137] Intl–C(2)[a138]		Global interpretability [a134,a135,a138], SHAP explanations [a137]	Clf[a137] Pred[a134–a136,a138]	
Gastroenterology and Hepatology [a139–a141]	Clinical Records [a139,a140] Lab Data [a139,a140] Imaging Data [a141]	NoCA–C(4),[a139] Intl–C(23),[a140] Intl–D(2)[a141]	FL[a141]	Global interpretability [a139]	Pred[a141] Risk/Stratify[a139,a140]	Web calculator [a140]
Other[a10,a142–a160]	Clinical Records [a10,a142–a146 a149,a151–a154,a156–a159] Lab Data[a10,a143 a151,a153,a156,a158,a159] Imaging Data [a10,a143,a146,a148,a149,a160] Survey Data [a143,a146,a147,a150,a153,a155] Genomic Data [a10] Digital Devices [a150]	Intl–C(2),[a142] CA–C(16),[a143] CA–C(pooled),[a144] CA–C(3),[a145] Intl–C(2200),[a10] NoCA(4),[a146] Intl–C(69),[a147] NoCA–C(2),[a148] Intl–C(pooled),[a149] NoCA–C(21),[a150] CA–C(5),[a151] NoCA–C(74),[a152] CA–C(2),[a153] NoCA–C(17),[a154] NoCA–C(15),[a155] Intl–C(3),[a156] Intl–C(10),[a157] NoCA–C(7700),[a158] Intl–C(>2),[a159] NoCA–D(8)[a160]	Risk-based anonymization[a10] OMOP CDM v5 across sites[a157] Distributed learning[a160]	Global interpretability [a143,a145,a146,a149,a151,a155,a156,a158,a159], Imaging saliency [CAM[a148]]	Clf[a145,a148,a159] Pred[a142,a143,a147,a149,a150,a160] Risk/Stratify[a151,a158] Enablement/Infra[a10,a144,a146,a152–a157]	Deployed private cloud (OpenStack)[a10] Shared code (GitHub) [a156,a159,a160]

Abbreviations: **CA** Canada-only; **Intl** international (including Canada); **NoCA** No Canadian data; **C** centralized; **D** decentralized; **FL** federated learning; **Clf** Classification; **Pred** Prediction; **Seg** Segmentation.

Table 2: Primary task catalogue by disease domain in multicenter AI collaborations with Canadian participation (2018–2024).

Domain	Primary tasks
Cancer & Oncology[a1, a7, a8, a11–a39]	**Clf** [mesothelioma subtype;[a11] overall survival (OS) + HPV status;[a12] primary cancer type (BM patterns);[a14] molecular subgroup;[a15] tumor detection + pediatric subtype;[a19] OS prognosis + ICI short/long survivors;[a20] MCED detection + site;[a22] radiogenomic subtype (BRAF vs others);[a30] HPV triage;[a32] benign vs malignant nodules;[a8] anastomotic leakage risk;[a34] Cancer detection (micro-US, biopsy targeting),[a35] multicancer subtype across sites,[a37] tumor subtype;[a38] Skin cancer detection (malignant vs benign) and multiclass lesion classification[a39]] **Pred** [LOS after oral cancer surgery;[a1] survival;[a12] CVD risk from calcifications;[a16] OS/DFS + surgery benefit;[a17] OS + local/regional control;[a21] local failure post-SBRT;[a23] BM progression after SRS;[a26] 2y OS + lifetime risk;[a27] PD-L1 + PFS stratification;[a28] PFS biomarker under ICI;[a29] PFS + OS (1st-line);[a31] LNM/recurrence after resection;[a33] nodule malignancy;[a8] csPCa detection + biopsy guidance[a36]] **Seg** [glioblastoma sub-compartment boundaries;[a7] PET tumor volume[a18]] **Risk/Stratify** [traumatic stress prevalence/severity + predictors;[a13] benefit-based groups;[a17] risk triage;[a21] outcomes + care-quality (STEMI cancer);[a24] surgery vs surveillance stratification;[a33] HPV reproducibility emphasis[a32]] **Enablement/Infra** [LHS analytics + dashboards + ML readiness[a25]]
Cardiovascular & Vascular Diseases[a4, a40–a64]	**Clf**[obstructive CAD;[a41, a43] HF case ID;[a45] Systolic/diastolic dysfunction;[a46] CAD rule-out;[a48] LVEDP class;[a53] Multi-disease ECG[a64]] **Pred**[30-day MALE/death;[a40] MACE w/ plaque+FFRCT;[a42] readmission AMI;[a50] Echo LV/RV quant;[a56] operative strategy (arch repair);[a57] CAD triage (CAC+CCTA);[a58] Warfarin dosing;[a4] AAA rupture;[a59] bleeding risk;[a60] OMI/culprit vessel;[a61] LV/RV dysfunction[a62]] **Seg**[MI segmentation (cine MRI)[a44]] **Risk/Stratify**[MACE hybrid MPI;[a47] AAA features ESUS;[a49] REFINE SPECT missing-data robustness;[a51] MACE risk (sex diffs);[a52] AAD mortality;[a54] Plaque nomograms;[a55] Aortic arch repair (death/stroke);[a57] Mortality (body composition CTAC)[a63]]
Neurological Disorders[a3, a65–a83]	**Clf**[PD vs controls;[a65] VNS responder vs non-responder;[a68] Benign vs malignant PNST;[a71] ALS vs controls;[a3] ADHD vs control; ASD vs control;[a78] PD vs controls (T1 MRI)[a81]] **Pred**[MCID at 6/12/24 mo;[a67] Fast vs slow RT;[a70] Brain age (regression);[a72] Progression / clinical change;[a75] Functional outcomes from network markers;[a77] Age/sex prediction;[a82] Chronic motor function;[a83] Motor-impairment severity proxy[a65]] **Seg**[Brain extraction / skull-strip (FLAIR);[a69] 10-class tissue segmentation (DWI)[a76]] **Risk/Stratify**[Structure–function links;[a66] Phenotyping / cohort stratification;[a73] Lesion–symptom mapping; sex-specific;[a74] Trial stratification / enrichment;[a75] Disease subtyping / progression;[a77] Cohort discrimination; risk signatures;[a79] Genetic association & fine-mapping;[a80] Normative deviation associations[a82]]
Psychiatric & Mental Health[a2, a5, a84–a99]	**Clf**[OCD vs controls;[a2] Bipolar vs controls;[a5] OCD vs HC; subgroup classifiers;[a87] MDD vs controls[a91]] **Pred**[Alcohol use;[a85] Depression onset (2- & 5-yr);[a86] Brain-age gap → ADT response;[a88] Lithium response;[a89] Differential remission across 5 monotherapies and 3 combinations;[a90] Future psychosis (CHR);[a93] rTMS response (≥20% PANSS-NS reduction);[a94] Binary remission + best drug recommendation;[a96] Predict non-adherence risk (≥ 12 sessions definition) to target peer-support[a99]] **Risk/Stratify**[cognitive ability mediates effects of genes & environment on PLEs;[a84] Phenotyping / staging (case–control + illness duration);[a92] Brain-age deviation ↔ cognition in MDD;[a95] Relationship quality;[a97] HCP mental health risk actors during COVID-19[a98]]
Respiratory & Lung Conditions[a100–a112]	**Clf**[Identify COPD;[a101] Normal vs abnormal LUS;[a102] Patient sex from CXR;[a103] physician-diagnosed asthma;[a104] Sliding-artifact detection;[a108] Wheeze vs normal;[a110] Multiclass disease classification[a111]] **Pred**[COPD admission & 30-day readmission[a107]] **Risk/Stratify**[Genetic variant ↔ IgE;[a100] Resource use & expenditures;[a105] Subtype discovery;[a106] Protein intake ↔ MV outcomes/weaning;[a109] Mortality risk stratification in ILD using CT-based ML score[a112]]
Kidney & Renal Diseases[a6, a113–a118]	**Clf**[Donor–recipient molecular compatibility[a6]] **Pred**[Short- & long-term graft failure, mortality, rejection;[a6] 1–5 yr kidney failure & all-cause mortality;[a113] CKD progression (40% eGFR decline or kidney failure, 1–5 yr);[a114] 30-day adverse outcomes[a117]] **Risk/Stratify**[Identify undertreated CKD pts;[a115] TG patient stratification & risk of graft loss;[a116] Prognosis for ≥40% eGFR decline or kidney failure (1–3 yr)[a118]]
Infectious Diseases [a119–a124]	**Pred**[Oxygen requirement, MV/death;[a119] CXR severity/trajectory;[a120] Reaction occurrence[a122]] **Risk/Stratify**[Immune response post-booster;[a121] Comparative effectiveness of CAP regimens[a123]] **Enablement/Infra**[Validate case-finding algorithms, screening strategy[a124]]
Critical Care[a125–a129]	**Clf**[Prolonged MV >14d;[a125] Delirium from notes;[a127] 30d mortality/AKI classification under scarce data[a129]] **Pred**[ICU mortality;[a126] Early ARDS mortality (≤24h);[a128] ICU/hospital LOS regression[a129]]
Surgical & Trauma Conditions[a9, a130–a133]	**Clf** [T-cell–mediated rejection;[a9] Molecular diagnosis of ABMR/TCMR (±injury);[a132] ABMR/TCMR diagnosis, ensemble report[a132]], **Pred** [graft loss;[a9] post-op transfusion need (THA)[a133]], **Risk/Stratify** [phenotype discovery of rejection/injury vs histology/DSA;[a130] frailty via CT muscle score for 30-day outcomes[a131]]
Reproductive, Pregnancy, and Neonatal Health[a134–a138]	**Clf**[vaginal birth after dinoprostone IoL[a137]] **Pred**[severe neonatal morbidity (<32 wks),[a134] postnatal gestational age estimation,[a135] semen concentration upgrade (post-varicocelectomy),[a138] AI-driven prognostic counseling (IUI/IVF conversion)[a136]]
Gastroenterology and Hepatology[a139–a141]	**Clf**[detect hepatic steatosis S0 vs. ≥S1[a141]] **Risk/Stratify** [ED-arrival triage for in-admission adverse outcomes;[a139] 30/90-day mortality risk aiding steroid/LT triage decisions[a140]]
Other[a10, a142–a160]	**Clf**[High vs low acuity (CTAS 1–3 vs 4–5) for remote triage;[a145] detect elbow/shoulder dislocation on X-rays;[a148] distinguish GCA (active/inactive) vs controls from plasma[a159]] **Pred**[Multi-task clinical prediction with few-shot label-efficiency;[a142] predict disease course/remission risk; subtype joint-involvement patterns;[a143] regress contact-avoidance, hygiene-maintenance, and policy-support scores;[a147] predict recurrence after Bankart repair;[a149] predict likelihood of wearable use in children;[a150] retinal-age regression; centralized vs FL vs TM[a160]] **Risk/Stratify** temporal-cluster detection to stratify deterioration risk;[a151] non-response risk and burden (HRU changes)[a158]] **Enablement/Infra** [Re-identification leakage from shared embeddings; anonymization comparison;[a144] Research-ready data platform to capture/anonymize/harmonize multi-source clinical data;[a10] Best practices for handling missingness in supervised ML;[a146] ICU spillover effects; counterfactual capacity-expansion simulations;[a152] Implement AE & lipid indicators, identify practice gaps, and support counseling;[a153] Real-world initiation of second-line antihyperglycemics by region/CVD risk;[a154] Measure prevalence/incidence at population scale;[a155] Compare diagnosis- vs lab-based labels; quantify cross-site variability;[a156] Second-line antihyperglycaemics—3/4-point MACE outcomes[a157]]

Abbreviations: **Clf** Classification; **Pred** Prediction; **Seg** Segmentation.

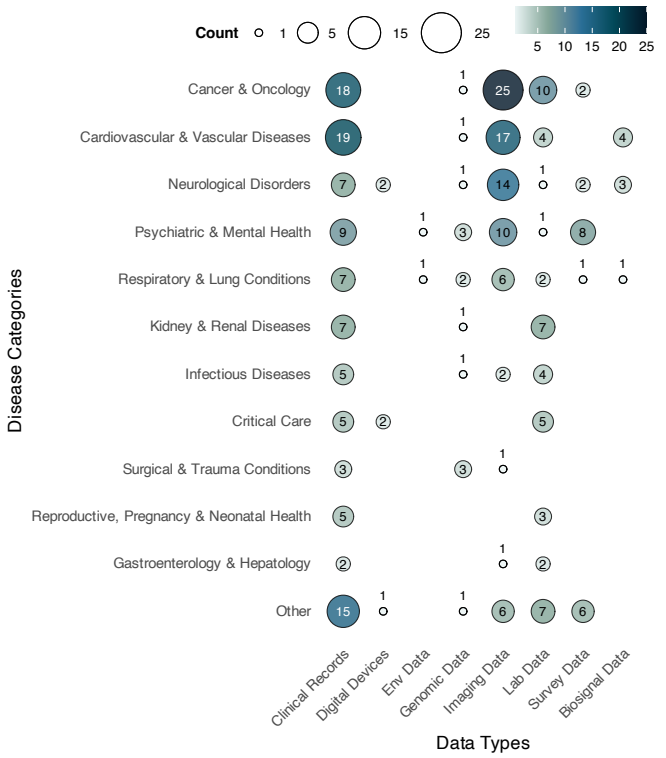


Figure 5: Bubble plot of disease categories by data modalities in Canadian multicenter data collection collaborations (160 studies). Bubble size and color indicate the number of studies, with larger and darker circles representing higher counts. Cancer & oncology and cardiovascular diseases dominate collaborations, with imaging and clinical records being the most frequently used data types.

Central Data Repository in Multicenter Collaborations across Canada

Centralized data repositories, where patient-level data from multiple sites are pooled into a single database, were used in the majority of multicenter collaborations in our review. Adhesion to the Findable, Accessible, Interoperable, and Reusable (FAIR) data sharing principles helped support cross-site data harmonization.[34] However, concentrating sensitive patient information in one location also raises privacy concerns, which studies addressed through measures like data de-identification, encryption, and strict access controls. As shown in Figure 1, 152 of the 160 multicenter collaborations (95%) in our dataset relied on a central data repository model, while only 8 studies (5%) used a decentralized approach. Of those 152 centrally coordinated studies, 31 were drew exclusively on Canadian patient-level data. These centralized collaborations led by Canadians offered information on common practices around ethics approvals, dataset sizes, and the use of advanced AI techniques. Most of these studies obtained clearance from multiple institutional Research Ethics Boards (REBs), and several retrospective analyses received informed consent waivers due to the use of de-identified or previously collected data. Some projects also implemented automated de-identification pipelines to facilitate compliance with privacy regulations during data aggregation.[35]

Dataset Scope and Ethics/REB Practices

Dataset sizes in the 32 exclusively Canadian multicenter studies varied widely. Eleven studies (34.4%) had fewer than 1,000 patient records, often focusing on rare diseases or pilot projects where data were inherently limited. Ten studies (31.3%) used medium-sized datasets of 1,000-10,000 patients, typically in clinical contexts such as breast cancer and heart failure. Four studies (12.5%) included about 10,000-100,000 patients, and another four studies (12.5%) analyzed between 100,000 and 1,000,000 records. Three studies (9.3%) examined cohorts that exceed one million patient entries, showing the ambition and resource requirements of nationwide data initiatives. This variation in the scope of the dataset reflects the known challenges in accessing very large medical datasets. It also illustrates that while some projects used extremely large 'big data' cohorts, many Canadian studies still worked with comparatively small sample sizes. The use of larger datasets aligns with the general expectation that more data can improve model performance, while the prevalence of smaller datasets confirms ongoing efforts to develop machine learning methods that remain effective with limited data.[36,37]

AI Methods, Explainability, and Deployment

A diverse array of AI and machine learning techniques was applied across collaborations using centralized data repositories (Table 3). Neural networks (36 studies, 23.7%) and random forests (35, 23.0%) were the most frequently used approaches, with neural networks valued for their adaptability to complex modalities such as imaging and free text, and random forests chosen for their robustness and strong performance on heterogeneous, tabular clinical data. Logistic regression (33, 21.7%) Continued to play a central role because it produces interpretable, well-calibrated risk estimates and has long been standard practice in clinical prediction and decision support.[38,39] Consistent with this, a large systematic review found that, across diverse clinical tasks, machine-learning models did not outperform logistic regression on discrimination or calibration.[39] Ensemble methods such as gradient boosting (30, 19.7%) and support vector machines (30, 19.7%) were also widely applied, particularly for structured datasets where predictive accuracy was prioritized. Regularized regression techniques (e.g., LASSO, Ridge, Elastic Net; 20, 13.2%) were frequently chosen to manage high-dimensional clinical variables, while survival models (19, 12.5%) addressed the longitudinal and time-to-event nature of many health outcomes. Dimensionality-reduction methods like PCA (15, 9.9%) supported exploratory analyses in high-dimensional omics or imaging settings. Decision trees (10, 6.6%), though less com-

Table 3: Frequently used AI techniques (>10 Studies) in multicenter Canadian studies with a central data repository

AI Technique(s)	#(%)
Neural Networks	36 (23.7%)
Random Forest	35 (23.0%)
Logistic Regression	33 (21.7%)
Gradient Boosting	30 (19.7%)
Support Vector Machines	30 (19.7%)
LASSO/Ridge/Elastic Net	20 (13.2%)
Survival Models	19 (12.5%)
Principal Component Analysis (PCA)	15 (9.9%)
Decision Trees/CART	10 (6.6%)

mon, provided simple and interpretable rules in clinical scenarios that required transparent decision-making. These patterns show that researchers tend to choose methods based on the problem at hand, aligning the model with the data type and the clinical decision it is meant to support. In practice, this often means prioritizing reliable and interpretable models that clinicians can trust, rather than using newer methods for novelty alone.

Beyond these established approaches, only two studies investigated foundation models. One study applied a foundation model for structured EHRs pretrained on 2.57 million patient records and reported that local adaptation matched gradient boosting performance while being 60-90% more sample-efficient.[a142] In the imaging domain, another study employed a histopathology foundation model with adversarial Fourier-based domain adaptation (AIDA) across four cancer types, which enhanced cross-site generalization compared to standard normalization approaches and generated interpretable heatmaps that corresponded to pathologist annotations.[a37]

Explainable artificial intelligence (XAI) strategies in these studies were grouped into three categories, including global feature importance (model-level rankings), local post hoc methods (instance-level attributions), and imaging saliency (visual maps of relevant regions). Despite the recognized importance of transparency, only 63 of 152 studies (41.4%) reported any XAI approach. Most relied on global methods (44, 28.9%), such as regression coefficients or feature importance scores (e.g., XGBoost gain, random forest impurity, permutation rankings). Local methods were infrequent, with only 10 studies (6.5%) applying post hoc instance-level explanations (e.g., SHAP, LIME) to generate patient-level attributions. Imaging saliency was implemented in 9 studies (5.9%), typically through heatmap-based approaches (e.g., CAM, Grad-CAM, SmoothGrad), and was often combined with expert validation (e.g., pathologists confirmed histologic regions highlighted in histopathology heatmaps[a37]). Several other studies used dimensionality reduction (e.g., PCA, t-SNE, UMAP) or performed calibration and robustness analyses (e.g., Brier score, feature stability), which can support transparency but fall short of post hoc XAI. These findings indicate that global methods predominate in structured/tabular data, saliency is the preferred approach in imaging, and local post hoc explanations remain underused.

Deployment patterns across the included multicenter studies emphasized incremental translational steps rather than full clinical integration. As summarized in the 'Interop/deploy' column of Table 1, the most common artifacts were code releases (11 studies, 7.2%; for example, a Venous Thromboembolism (VTE) risk analysis study made its modeling scripts publicly available[a60]) and web calculators (4 studies, 2.6%; for example, an AI-generated 90-day survival model for alcoholic hepatitis was deployed as a live Shiny app[a140]). More advanced efforts included containerized services, such as Docker-packaged models with REST APIs,[a6,a25] and one infectious disease collaboration that distributed pretrained models through NVIDIA GPU Cloud (NGC) via Clara Train,[a119] signaled initial moves toward reproducible, infrastructure-aware deployment. Beyond these point solutions, a smaller number of projects emphasized interoperability standards, particularly the adoption of OMOP CDM across sites, to support harmonization and scalability for downstream applications.[a50,a142,a156,a157]

Decentralized Data Repository in Multicenter Collaborations

A decentralized data repository is a system where data stay at multiple originating sites rather than being pooled centrally, which can improve scalability, privacy, and security. Decentralized approaches are integral to federated learning (FL), which enables collaborative model development across institutions without centralizing sensitive data in regulated settings such as healthcare.[21] In Canadian healthcare research, adoption of decentralized data repositories remains in its early stages. Only 8 of 160 multicenter collaborations (5%) used decentralized designs, covering three patterns: (i) FL (6 studies[a6,a7,a64,a119,a141,a160]), (ii) traveling models (TM) (2 studies,[a81,a160] the latter of which also evaluates FL), and (iii) distributed classical ML.[a12] In FL, sites train locally and periodically send model updates to a coordinator, which preserves data locality and often matching pooled-data accuracy, but this setup requires round-based synchronization, adequate bandwidth, and careful handling of non-IID client distributions, such as class imbalance and scanner/site effects. TM removes the aggregator and moves a single model sequentially across sites, reducing orchestration and bandwidth demands, but introducing order effects and slower global state refresh than FL. Distributed classical ML keeps feature engineering and model fitting simple and interpretable, which allows clinicians to audit parameters, but comes at the cost of lower representational capacity and stronger reliance on harmonized preprocessing.

Although usage is limited, these studies show the potential of decentralized strategies for rare conditions where no single center has enough patients. Table 4 compares key aspects of the eight studies that used decentralized data sharing, including federated and distributed learning. In the federated learning projects we identified, FL enabled teams to train models jointly without sharing raw data and improved generalizability by learning from more diverse populations. We also observed a gap in long-term sustainability plans, as none of the studies described how the federated network would be maintained over time. None of the FL studies reported model explainability methods. We identified only one collaboration conducted entirely within Canada.[a64]

Interoperability issues from siloed provincial health information systems limit data sharing and slow the adoption of federated networks in Canadian healthcare.[40] Even without a purely domestic federated network, Canadian researchers take part in international decentralized collaborations. Canada therefore contributes to multinational federated initiatives that enable secure large-scale analytics focused on rare-disease research. In practice, each institution retains control over its data and training is coordinated by sharing only model updates or aggregated parameters, which allows learning across sites without moving sensitive patient information outside the source institution.

Table 4: Comparison of multicenter decentralized data repositories in Canadian healthcare research

	Bogowicz 2020[a12]	Dayan 2021[a119]	Schapranow 2023[a6]	Bakas 2022[a7]	Souza 2024[a81]	Agrawal 2024[a64]	Qi 2024[a141]	Nielsen 2024[a160]
Clinical domain / task	Head & Neck cancer: HPV & 2-yr OS	COVID-19: O_2 need (24/72h)	Kidney transplant CPMs (risk)	Glioblastoma: tumor boundaries (ET/TC/WT)	Parkinson's disease classification	ECG multi-label diagnosis	Hepatic steatosis detection	Retinal age prediction (RAG)
Participating sites (scope)	6 cohorts (multi-institution)	20 institutions (global)	Germany + Canada (major centres)	71 sites across 6 continents	83 centres (global; many small)	7 hospitals (province-wide)	2 centres (private + public; simulated clients)	UK Biobank (up to 2,400 simulated clients) + BRSET
Population / sample size	1,174 patients	~16,000 ED presentations	~8,000 retrospective cases	6,314 cases	1,817 scans	1,565,849 ECGs / 243,128 patients	n = 208 (153 + 55); 2,080 images	16,630 (UKB) + 958 (BRSET) images
Modality	CT radiomics	CXR + EMR (vitals, labs)	Clinical; HLA typing + matching; labs; pathology	mpMRI	3D brain imaging	12-lead ECG time-series	Ultrasound (B-mode) + biopsy steatosis grades	Fundus photos via RETFound features
Distributed model	Distributed feature selection + logistic regression	Federated learning (FedAvg) "EXAM"	Federated learning on private blockchain	Federated consensus segmentation	Traveling Model (TM); sequential site updates	FL with Differential Privacy (DP-SGD)	Simulated FL (four algorithms incl. FedAvg)	Foundation-model-driven deep learning; compare FL vs TM
Infrastructure / deployment	EuroCAT / DistriM-style distributed setup	NVIDIA FL framework; multi-site federation	Private blockchain; consortium pipelines	Global aggregation server; international consortium	Sequential TM; no central aggregation	Multi-hospital secure servers; site-wise eval	Flower-based simulation; partition strategies	8-bit features for efficiency; low-resource friendly
Reported performance (high-level)	Distributed ≈ centralized (no AUC drop; identical LR coefficients)	AUC > 0.92; improved generalizability vs single-site	N/A (protocol)	DSC ↑ vs public model — local: +27% (ET), +33% (TC), +16% (WT); OOS: +15% (ET), +27% (TC), +16% (WT)	TM AUROC ≈ 83% vs 80% centralized	FL (81.03%) comparable to pooled (82.07%); benefits for small-data sites	FedAvg AUC 0.93; central 0.90; single-site 0.83; imbalance hurts	MAE ~3.6 years; TM converges faster; FL/TM ≈ centralized
Explainability methods	Transparent LR coefficients	-	-	-	-	-	-	-
External / out-of-sample validation	Internal validation across cohorts; no dedicated held-out institutions	Independent external hospitals (incl. largest test site); multi-site evaluation	-	6 held-out sites (OOS) for generalization	Real-centre TM; compared to centralized baseline; no separate external set	Internal across 7 hospitals; no external institutions	Held-out test split; no external centres (simulated clients)	External BRSET evaluation in addition to UK Biobank
Client heterogeneity & non-IID handling	Multi-cohort radiomics; no explicit non-IID method	Inter-site EMR/CXR heterogeneity; assessed via multi-site training & external testing (no special non-IID algorithm)	Cross-national heterogeneity anticipated; harmonization/QC pipelines specified (protocol)	Scanner/site variability addressed by standardized mpMRI preprocessing; OOS sites held out	Many small, heterogeneous real centres; TM chosen for non-IID/minimal-data settings	Hospital-wise label prevalence differs; FL competitive with pooled, esp. for low-data sites	Explicit non-IID experiments (class imbalance & partitioning) quantify performance drop	Simulated client heterogeneity; TM converges with fewer local updates than FL

Abbreviations: **ED** = emergency department; **EMR** = electronic medical record; **CXR** = chest X-ray; **mpMRI** = multi-parametric MRI; **ET** = enhancing tumor; **TC** = tumor core; **WT** = whole tumor; **OOS** = out-of-sample (held-out sites); **TM** = traveling model; **FL** = federated learning; **DP-SGD** = differentially private SGD; **LR** = logistic regression; **AUC** = area under the ROC curve; **MAE** = mean absolute error; **CI** = confidence interval; **UKB** = UK Biobank; **BRSET** = Brazilian Multilabel Ophthalmological Dataset.

Discussion

Our scoping review reveals a clear contrast between Canada's advancing AI work and its limited data sharing practices. AI models are now more powerful than ever, capable of learning from text, images, genomics, and beyond, with documented gains in precision medicine. However, Canadian researchers face real hurdles in accessing and managing the diverse patient data needed to put this potential to work. This study found that despite available tools such as multimodal models and federated learning, Canadian teams rarely adopt these approaches in practice. Most collaborations still centralize data or analyze datasets held outside Canada.

Impact of patient-level Data on Precision Medicine

Across the studies we reviewed, patient-level data were essential to precision medicine, enabling more individualized care and deeper insight. In this scoping review, multicenter collaborations involving Canada primarily used imaging modalities, mainly in cancer and cardiovascular studies, but we also observed a growing inclusion of genomic data and patient-reported outcomes. This shift toward multimodality indicates that precision medicine research is expanding beyond biological signals to include behavioral and social determinants of health. It signals a move toward a more comprehensive, whole-person approach to personalized care. One clear example of the power of integrating patient-level data comes from the United Kingdom's 100,000 Genomes Project. This national initiative successfully embedded whole-genome sequencing into routine care, with participants consenting to link their genomic data to longitudinal clinical records in a secure environment that protected privacy.[41–43] As a result, the program has provided new diagnoses for thousands of patients with rare diseases and cancer, demonstrating tangible clinical benefits from combining genomic information with individual health data. Similarly, the United States' All of Us program has already enrolled over 500,000 participants, linking EHRs with genetic profiles to enable nationwide precision medicine studies.[44] Open resources like the UK Biobank, a cohort of 500,000 individuals with harmonized health data, have enabled discoveries worldwide by providing a large and standardized dataset.[45] These successes highlight that assembling rich, population-scale patient-level datasets can significantly accelerate biomedical innovation.

In this context, our review finds that Canada's research ecosystem, which is heavily international, struggles to build comparable domestic resources. Imaging was the most common data type in Canadian AI collaborations, appearing in just over 22% of studies. Radiology's early adoption of standardized formats such as DICOM makes images relatively easy to share across sites, and many collaborative AI projects in Canada were image-based.

In contrast, genomics and biosignals remain underused. Integrating them is harder because genomic data are analytically complex and computationally intensive, and EEG/ECG signals are often inconsistent and noisy. However, these modalities are crucial for models that account for a patient's condition across modalities. By 2024 we observed an increase in collaborations that include genomic information, indicating movement toward a multimodal approach as data-management tools improve. Canada still lags on this front. Without large, unified national datasets, investigators often turn to non-Canadian datasets to achieve the heterogeneity needed for robust models, a pattern reflected in our finding that nearly two-thirds of collaborations were international.

International Collaboration vs. Domestic Fragmentation

A key finding is that Canadian-led AI collaborations mostly involve international partners. This global orientation can facilitate knowledge exchange and give teams access to larger cohorts and specialized resources abroad, but it also exposes limited domestic capacity. In our dataset, almost two-thirds of collaborations with Canadian involvement were international. Only about one third were purely domestic, and among 160 multicenter studies, only 32 (20%) pooled data across Canadian sites. This imbalance indicates that interprovincial barriers, from misaligned privacy rules to technical incompatibilities, still hamper large-scale data pooling within Canada.

International precedents show what is possible when robust national resources are in place. The United States' *All of Us* program has enrolled more than 500,000 diverse participants, linking EHRs with genomic analyses to enable nationwide precision medicine research.[44] The United Kingdom's Biobank, an open-access cohort of 500,000 individuals with standardized health data, has catalyzed thousands of studies by providing a broadly generalizable research dataset.[45] Canada currently lacks an equivalent platform.

We found that 64.2% of collaborations involved international partners, which suggests that Canadian scientists often rely on foreign data. This dependency is a strategic risk because if most models are trained on non-Canadian populations, they may not generalize well to Canadian settings. For example, a diabetic retinopathy model trained largely on images from abroad could underperform on Canadian patients if demographic or environmental risk factors differ. Many studies that could plausibly have been done entirely within Canada, had data sharing been easier, relied on international partners to reach the needed scale or diversity. In practice, it is often easier for a Toronto-based researcher to work with a Boston collaborator under a shared framework than to navigate separate agreements with colleagues in Montreal or Vancouver. It can be easier to collaborate across an ocean than across provincial lines.

Our findings point to an urgent need to strengthen intra-Canadian research linkages. Building robust national data resources and networks would enable Canadian scientists to develop AI models and precision medicine insights tailored to Canada's population and healthcare context. It would also help ensure that Canadian patients benefit directly from research conducted in Canada. Some steps are underway. Legislative efforts such as Bill C-72 aim to promote secure and interoperable health data exchange across provinces. Pan-Canadian initiatives are beginning to establish common standards and shared infrastructure to reduce fragmentation. For example, the

Marathon of Hope Cancer Centers Network is linking major cancer centers to share clinical and genomic data, with the goal of creating a 'gold cohort' of 15,000 patients for precision oncology. Success will depend not only on technology and policy, but also on a culture of trust and collaboration among Canadian institutions. International partnerships will remain important, but strengthening domestic data sharing and collaboration is the path from supporting others' advances to leading transformative work in health AI.

The Potential of LLMs to Enhance Healthcare Research

We also see the rise of large language models (LLMs), especially multimodal models, as a promising tool in healthcare research. As clinical datasets expand to include text, images, signals, and more, advanced LLMs can integrate these heterogeneous inputs and reveal patterns that single-modality models miss. Recent work illustrates this potential. Google's Med-PaLM 2 reached expert-level scores on US medical licensing exam questions, suggesting the feasibility of high-quality clinical knowledge support.[46] GatorTron, trained on more than 90 billion words of de-identified electronic health records, achieved strong accuracy on tasks such as medical concept extraction and relation identification.[47] Vision-language systems are also advancing. PathAlign, trained on more than 350,000 pathology images paired with text, generated detailed diagnostic reports that pathologists judged about 78% acceptable, preserving key clinical details while automating a labor-intensive step.[48] In radiology, RaDialog integrates image analysis with an LLM-based question-answering interface; it produces complete chest X-ray reports and interactively answers clinicians' questions about the findings.[49] These examples show how multimodal AI can augment expert workflows and improve efficiency and consistency in radiology, pathology, and beyond.

As LLMs advance, their ethical and safety implications require attention. Patient privacy is central. Large models can unintentionally memorize and reproduce identifiable details when trained on sensitive health data, highlighting the need for safeguards.[50] Techniques such as differential privacy and careful filtering of training data can mitigate this risk. Accuracy and accountability also matter. Models like Med-PaLM 2 and RaDialog are impressive, but they can produce incorrect or nonsensical outputs and can miss clinical context. For now, LLMs should serve as assistive tools for clinicians rather than autonomous decision makers.

Privacy-Preserving Collaboration

The case studies of federated learning in our review (Table 4) show how privacy-preserving techniques can enable collaboration that is otherwise limited by data-sharing restrictions. For example, Bogowicz et al. trained a head-and-neck cancer outcome model between institutions in different countries without pooling data, by periodically aggregating model updates from each site.[a12] Dayan et al. likewise predicted COVID-19 severity across 20 hospitals on multiple continents with high accuracy (AUC 0.92 for mortality risk) while each hospital retained full control of its own data.[a119] These studies indicate that federated learning can match the performance of models trained on centralized data, even under strict local privacy rules.

Traditional multicenter research relied on de-identified data warehouses. Platforms such as i2b2 and Canada's GEMINI network showed that hospitals could contribute de-identified patient records to a central repository for analysis across many sites.[51,52] De-identification reduces risk but does not guarantee privacy. For example, Sweeney re-identified a public figure's health record by linking an 'anonymous' medical dataset with voter records using only birth date and postal code.[53] This vulnerability spurred the development of more advanced privacy-preserving techniques. Differential privacy adds statistical noise to the results of the query to protect individuals while preserving aggregate patterns.[54] Synthetic data generation creates artificial records that mimic real datasets without exposing individual data points.[55] Cryptographic methods such as homomorphic encryption and secure multi party computation enable computation on encrypted data or joint analyses without revealing raw data.[56,57]

The federated learning studies in our review used some of these techniques in practice. Schapranow et al. augmented a two-site kidney transplant federation with a blockchain ledger to audit data exchanges,[a6] and Pati et al. incorporated rare-case data augmentation to ensure that small hospitals' data still influenced a global brain tumor model.[a7] All projects encrypted communications between sites and some normalized or stratified data locally to mitigate inter-site variability. The overarching lesson is that privacy-preserving collaboration is feasible, but it requires careful planning and technical support. Coordinating data schemas, secure communication protocols, and governance agreements across many institutions is non-trivial. As Canada and other countries invest in learning health systems, embedding privacy-by-design into new health data initiatives will be essential.

Maintaining Model Performance and Addressing Challenges with Data Heterogeneity and Class Imbalance

The federated approaches we examined generally achieved performance comparable to traditional centralized training. For example, Qi et al. reported that their federated liver ultrasound model's AUC ranged from 0.85 to 0.93 on different test sets, closely matching the performance of a model trained on combined data.[58] Pati et al. likewise observed significant improvements in detecting glioblastoma tumor boundaries when training on a federation of 71 sites, relative to using a much smaller public dataset.[a7] These results suggest that, given sufficient quality and diversity of data at each site, federated models need not sacrifice accuracy for privacy.

Nevertheless, maintaining high model performance across disparate datasets poses challenges. A key factor is data heterogeneity. If one hospital's imaging protocols or patient demographics differ greatly from another's, a single global model may not fit all distributions well. In Qi et al.'s study, for instance, the model's performance dipped, with AUC dropping as low as 0.34, when a local site's data distribution diverged

markedly from the global distribution. This highlights the importance of strategies like data harmonization and, in some cases, localized model fine-tuning to handle site-specific quirks. It also highlights a practical advantage of centralized datasets, as data can be merged and cleaned into a more uniform whole before modeling, whereas in federated setups the inconsistencies must be managed algorithmically.

Class imbalance is another common issue. In many medical datasets, certain conditions or outcomes are underrepresented, which can skew training and reduce a model's ability to generalize. In a multicenter scenario, one hospital might contribute most of the rare cases while another contributes mostly controls. Without correction, the global model may be biased toward the majority class and underperform on the minority class. We saw this in a rare-cancer federation by Schapranow et al., where variability in predictions was observed due to combining heterogeneous datasets without proportionate class representation.[a6] To address such issues, researchers employed techniques like oversampling of minority cases and cost-sensitive learning to ensure that federated models remained fair and sensitive even when some outcomes were relatively scarce. These remedies can improve a model's reliability and equity, but they require careful implementation and validation in distributed settings.

Future Directions for Collaborative AI in Healthcare

While federated learning and similar approaches show promise, several areas need further work to realize the benefits of collaborative AI in healthcare. A priority is to standardize federated learning protocols across institutions and medical domains. Projects such as ENIGMA-Epilepsy and NephroCAGE show that each clinical domain often requires customized models and workflows, which can slow adoption if every collaboration starts from scratch.[59,a6] Tools like NVIDIA FLARE offer adaptable frameworks that support domain-specific requirements while maintaining a core of standardized processes, which helps streamle the deployment of federated learning in both research and clinical practice.[60] Adopting such frameworks can reduce the technical burden on participating sites and make federated studies more efficient and repeatable.

Emerging applications present opportunities for federated approaches. Continuous patient monitoring for chronic disease management, real-time predictive modeling in critical care units, and adaptive clinical decision support tools all are promising use cases for federated learning in near real-time settings. Developing frameworks that support continuous learning, updating models as new data become available, could substantially change the way AI is used at the bedside, enabling models that evolve to incoming patient data to support real-time care.

Beyond technology, inclusive research environments are essential. Many healthcare institutions, especially smaller or under-resourced ones, struggle with the infrastructure and expertise needed to integrate AI into practice.[61,62] Issues such as incompatible data systems, a shortage of technical staff, and uncertainty about data governance can widen the gap between well-resourced centers and others.[63] Without intervention, the benefits of AI, from enhanced diagnostics to personalized treatments, may accrue mainly to large hospitals, worsening healthcare disparities as smaller providers fall behind. To prevent this outcome, it is important to develop user-friendly platforms (e.g., low-code or no-code tools) that allow clinicians and other non-technical users to participate in AI development and deployment. Notably, national efforts like Canada's interoperability roadmap aim to reduce these barriers. Ultimately, lowering entry barriers to AI adoption, encouraging broader collaboration, and simplifying the use of privacy-preserving technologies will help ensure that AI-driven improvements reach the full healthcare system, not only the most resourced institutions.

Limitations

Although scoping reviews often use multiple databases, PRISMA-ScR permits a single database when justified.[22] We limited our search to PubMed after a pilot assessment against IEEE Xplore, ACM Digital Library, Scopus, and Web of Science showed comparable coverage for our eligibility criteria. PubMed indexes more than 30,000 biomedical and clinical journals and incorporates open-access and non-MEDLINE content via PubMed Central.[64] Its structured records and programmatic interface (E-Utilities) enabled efficient, scalable extraction with our LLM pipeline.[65] This choice may miss AI studies published primarily in engineering venues; however, given our focus on patient-level data in health contexts and the need for reliable programmatic access, PubMed was the most appropriate and practical source at the time.

Second, our identification of Canadian collaborations and of studies using patient-level data relied on an automated pipeline, validated on a random sample of 750 papers. Validation showed that roughly 7% of papers labeled as using patient-level data were false positives on manual review. This likely arose from reliance on keywords in titles and abstracts that sometimes misrepresented a study' data level. While a 7% noise rate is acceptable for high-level trends in a large dataset, some individual studies were misclassified. Determining whether a collaboration was international or domestic based only on author affiliations is imperfect. For example, if a researcher held dual appointments in Canada and the United States but listed only the Canadian affiliation, our algorithm would tag the study as domestic even if the work involved international partnership.

Another limitation is inconsistent terminology in the literature, which we normalized for analysis. Different authors often used varying terms for the same concept (e.g., 'CT' vs. 'Computed Tomography', or 'electroencephalography' vs. 'EEG'). We mapped synonyms to a common vocabulary and grouped related items, though some nuances may have been lost. Our hierarchical clustering is only as good as the input matrix; if grouping or coding missed subtle distinctions, the resulting patterns would be affected.

Finally, we used a large language model (i.e., LLaMA 3) to assist with feature extraction from study texts. This enabled scale, but LLM outputs can be inconsistent and sensitive to presentation. Ambiguous or poorly structured abstracts may have led to missed or misinterpreted details about collaboration or

data usage. We incorporated human oversight and iterative refinement to improve accuracy, but we cannot rule out that certain nuances were not fully captured.

Despite these limitations, the major trends and observations in this review appear robust and provide a useful overview of collaborative medical AI research in Canada. Without intervention, Canadian researchers may remain data contributors to others' studies rather than leaders of homegrown precision medicine work.

Methods

Search strategy and selection criteria

Our search strategy began with a pilot assessment of multiple databases to identify the most comprehensive and technically feasible source for large-scale automated analysis. We evaluated PubMed, IEEE Xplore, ACM Digital Library, Scopus, and Web of Science for their coverage of AI-driven healthcare research involving patient-level data. This assessment showed that publications retrieved from the non-PubMed databases were fully covered within PubMed for our defined eligibility criteria. In addition, PubMed offers broad coverage of high-quality biomedical literature and provides structured XML records[66] that facilitate large-scale programmatic data extraction. These factors, combined with the technical requirements of implementing our automated pipeline, led us to focus our main search and analysis on PubMed. Consequently, all figures and results in this manuscript report the PubMed-derived dataset. A comprehensive electronic search was performed for publications from 2018 up to February 2025. This time frame captures a period of accelerating interest in machine learning applications for patient data. The search strategy combined keywords related to artificial intelligence, machine learning, and healthcare to retrieve relevant studies (the full query string and search parameters are provided in Supplementary Note 1). No restrictions were placed on study design within these parameters.

We included studies published in English that reported on AI-driven health research, involved at least one author with a Canadian institutional affiliation, and utilized patient-level data. We excluded articles that were review papers, opinion pieces (e.g., editorials or letters), or retracted publications. Studies relying solely on synthetic data, openly available public datasets, simulated data, or animal data were also excluded. These inclusion and exclusion criteria were designed to capture original research involving real human patient data and active collaboration, while filtering out literature that did not directly contribute empirical findings on the sharing of patient-level data. No formal critical appraisal of individual sources of evidence was performed. Consistent with scoping review methodology, all studies meeting the inclusion criteria were included regardless of their methodological quality. While the review did not assess study quality for inclusion decisions, the synthesis considers methodological factors that may have influenced the reported findings and their interpretation.

Study selection process

All references retrieved from the search were exported for screening. We implemented a two-phase screening process to identify articles meeting our eligibility criteria (Figure 6). In the first phase, we filtered records by author affiliation to find studies with Canadian contributions. This was accomplished through a combination of automated string matching for Canadian institution names and a large language model (i.e., LLaMA 3[67]) to resolve ambiguous cases, ensuring that at least one author of each included study was affiliated with a Canadian institution (details of the affiliation detection algorithm are provided in Supplementary Note 1.1). In the second phase, we screened for the presence of patient-level data in each study. We applied an LLM-based pipeline (i.e., a 7-billion-parameter model integrated via LangChain[68]) to parse titles and abstracts, and full texts when necessary, for indications that the research used individual-level patient data. This pipeline flagged and retained studies explicitly involving patient-level clinical data and excluded those that only utilized fully de-identified, aggregate, or open-access datasets (e.g., studies exclusively based on public databases like MIMIC-III[69] were filtered out). After these two filtering steps, only articles that had both a Canadian-affiliated author and a use of patient-level data remained for inclusion.

To verify the accuracy of the automated selection, we conducted a validation check on a random subset of the screened articles. Three independent reviewers with backgrounds in medical AI manually assessed 750 records that were randomly selected from different stages of screening and compared them with the pipeline's inclusion decisions. The agreement between the LLM pipeline and human judgment was substantial, with a false-positive inclusion rate of approximately 7%. This indicates that only a small proportion of articles were incorrectly flagged as meeting the criteria and were subsequently excluded in later stages of screening. Given the large initial sample size, this error rate was considered acceptable for the purposes of this study.

Features extraction and data analysis

We extracted relevant data from each study using a structured charting approach supported by natural language processing tools (Figure 6). Instead of relying solely on manual data extraction forms, we developed a custom pipeline based on retrieval-augmented generation (RAG)[70] to automate the data charting. This pipeline incorporated multiple advanced LLMs, including LLaMA 3, LLaMA 3.1, and LLaVA,[71] implemented within the LangChain framework[68] to identify and record key study characteristics from articles. In addition to text, the pipeline was capable of parsing tables and images extracted from the articles, allowing the capture of structured and visual information that is often overlooked in automated reviews. While this capability is a relatively novel aspect of the approach, it was applied in a targeted and pragmatic manner to enrich the dataset without overstating its role. Prior to full deployment, the extraction procedure was pilot-tested and iteratively refined using a subset of articles to ensure it captured

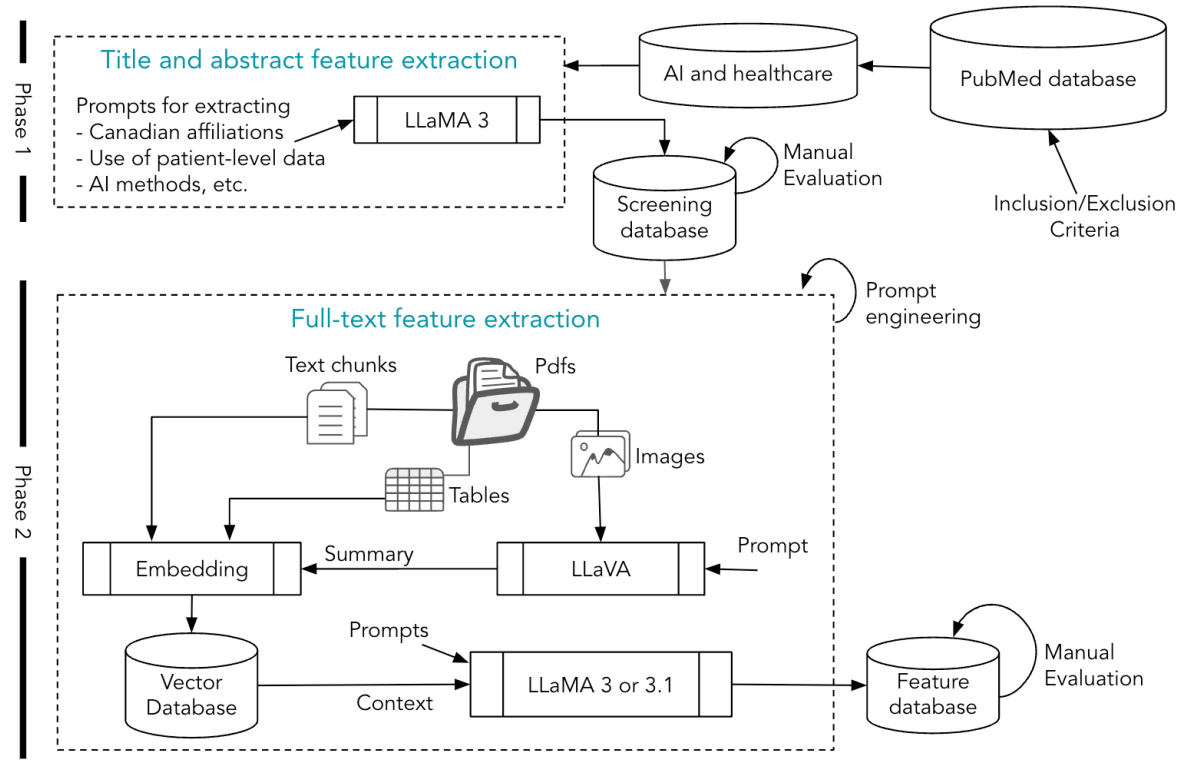


Figure 6: Visualization of the two-phase article selection and refinement process. The first phase applies automated filtering techniques to identify Canadian-affiliated authors and detect patient-level data from a set of PubMed articles. The second phase involves feature extraction using advanced models to capture key research features.

the desired information consistently. We then conducted a formal validation of the charting process with two reviewers independently examining the outputs for a random sample of 600 included studies, comparing the LLM-extracted data points to those obtained by manual review. Discrepancies were analyzed and used to adjust the pipeline (e.g., refining prompt templates or adding rules for edge cases), thereby improving the accuracy and reliability of the automated extraction. The feature extraction pipeline and tools used are described in the Supplementary Note 2.2.

For each included study, we charted a range of data items to facilitate analysis. Key variables extracted included bibliographic information (i.e., publication year and venue) and study characteristics relevant to collaboration and data use. We recorded each study's collaboration scope, categorizing it as *domestic* if all collaborating authors were from Canadian institutions or *international* if at least one collaborator was affiliated with an institution outside Canada. We noted the number of participating sites and whether the research involved a single-site dataset or a multicenter data collection effort. For multicenter studies, we further documented whether Canadian patient data were included in the dataset or if the Canadian collaborators worked only with data from non-Canadian sources. We also captured the modality of patient-level data used in each study, such as clinical imaging (e.g., MRI, CT, fMRI), electronic health records, biosignals (e.g., EEG, ECG), genomic or molecular data, and other modalities as applicable. Furthermore, we identified any mention of privacy-preserving or decentralized data sharing methods employed by the investigators. In particular, we flagged studies that utilized techniques like federated learning, secure multi-party computation, or differential privacy to enable multi-site collaboration without direct data pooling. All data item definitions were established a priori and built into the LLM extraction prompts to ensure uniformity in how information was captured across sources.

To summarize the charted data and address the review questions, we used descriptive statistics. The cleaned dataset (see Supplementary Note 2.3 for preprocessing details) was analyzed to determine frequencies and proportions for key features, such as the percentage of studies involving international collaboration or the distribution of data modalities across projects. Temporal trends were examined to observe changes in collaboration patterns over the 2018 to 2024 period. We visualized the geographic distribution of international collaborations by mapping the countries of partner institutions and ranking the most frequent collaborator countries outside Canada. To explore relationships between study characteristics, we performed exploratory clustering and cross-tabulation analyses (i.e., grouping studies by data modality and clinical domain to identify common patterns). Finally, we conducted a qualitative synthesis by reviewing relevant contextual information from the full texts of included studies. This qualitative review focused on common challenges and enablers of data sharing (e.g., requirements for multiple ethics board approvals, issues of data standardization, or governance frameworks mentioned) to complement the quantitative findings.

Funding

This research is supported by the Data Sciences Institute at the University of Toronto (DSI-PDFY3R1P31), and the Artificial Intelligence for Public Health (AI4PH) Trainee Award from the Canadian Institutes of Health Research (CIHR).

Acknowledgements

We acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Canada Research Chairs Program and Discovery Grant (RGPIN-2025-07037).

Author contributions

O.J. led methodology, software implementation, validation, formal analysis, data curation, and visualization, and drafted and revised the manuscript. Q.Z. and R.R. developed software, conducted formal analyses and investigations, curated data, and drafted the manuscript. M.N. contributed to methodology, drafting and revising the manuscript, and supervision. A.S. and B.F. reviewed and edited the manuscript. Z.S. conceived the study, led conceptualization and methodology, provided resources, supervised the work, secured funding, contributed to visualization, and drafted and revised the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

[1] Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).

[2] Venturini, S., Sikdar, S., Rinaldi, F., Tudisco, F. & Fortunato, S. Collaboration and topic switches in science. *Scientific Reports* **14** (2024).

[3] Lee, S. & Bozeman, B. The impact of research collaboration on scientific productivity. *Social studies of science* **35**, 673–702 (2005).

[4] Qin, J., Lancaster, F. W. & Allen, B. Types and levels of collaboration in interdisciplinary research in the sciences. *Journal of the American Society for Information Science* **48**, 893–916 (1997).

[5] Canadian Institutes of Health Research. Cihr strategic plan 2021–2031: Investing in the long-term health of canadians. https://cihr-irsc.gc.ca/e/52331.html (2023). Accessed May 30, 2025.

[6] Kulasegaram, K. *et al.* Data sharing and big data in health professions education: Ottawa consensus statement and recommendations for scholarship. *Medical Teacher* **46**, 471–485 (2024).

[7] Flanagin, A., Curfman, G. & Bibbins-Domingo, K. Data sharing and the growth of medical knowledge. *JAMA* **328**, 2398 (2022).

[8] Tamuhla, T., Lulamba, E. T., Mutemaringa, T. & Tiffin, N. Multiple modes of data sharing can facilitate secondary use of sensitive health data for research. *BMJ Global Health* **8** (2023).

[9] Tucker, K. *et al.* Protecting patient privacy when sharing patient-level data from clinical trials. *BMC Medical Research Methodology* **16** (2016).

[10] He, Y. *et al.* Foundation model for advancing healthcare: Challenges, opportunities, and future directions (2024). URL https://arxiv.org/abs/2404.03264. 2404.03264.

[11] Johnson, K. B. *et al.* Precision medicine, ai, and the future of personalized health care. *Clinical and Translational Science* **14**, 86–93 (2020).

[12] Alowais, S. A. *et al.* Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Medical Education* **23** (2023).

[13] Moor, M. *et al.* Foundation models for generalist medical artificial intelligence. *Nature* **616**, 259–265 (2023).

[14] Li, C. *et al.* Llava-med: Training a large language-and-vision assistant for biomedicine in one day. *Advances in Neural Information Processing Systems* **36**, 28541–28564 (2023).

[15] Acosta, J. N., Falcone, G. J., Rajpurkar, P. & Topol, E. J. Multimodal biomedical ai. *Nature medicine* **28**, 1773–1784 (2022).

[16] Tu, T. *et al.* Towards generalist biomedical ai. *Nejm Ai* **1**, AIoa2300138 (2024).

[17] Parliament of Canada. Bill C-72: An Act respecting the interoperability of health information technology and to prohibit data blocking by health information technology vendors (2024). URL https://publications.gc.ca/collections/collection_2024/parl/X1-2-441-72-eng.pdf.

[18] Leavy, P. *Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches* (Guilford Publications, 2022).

[19] Baxter, G. & Sommerville, I. Socio-technical systems: From design methods to systems engineering. *Interacting with computers* **23**, 4–17 (2011).

[20] Mumford, E. The story of socio-technical design: Reflections on its successes, failures and potential. *Information systems journal* **16**, 317–342 (2006).

[21] Yang, Q., Liu, Y., Chen, T. & Tong, Y. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* **10**, 1–19 (2019).

[22] Tricco, A. C. *et al.* Prisma extension for scoping reviews (prisma-scr): checklist and explanation. *Annals of internal medicine* **169**, 467–473 (2018).

[23] Peters, M. D. *et al.* Guidance for conducting systematic scoping reviews. *International journal of evidence-based healthcare* **13**, 141–146 (2015).

[24] Arksey, H. & O'Malley, L. Scoping studies: towards a methodological framework. *International journal of social research methodology* **8**, 19–32 (2005).

[25] Levac, D., Colquhoun, H. & O'Brien, K. K. Scoping studies: advancing the methodology. *Implementation science* **5**, 69 (2010).

[26] Public Health Agency of Canada, Pan-Canadian Health Data Strategy Expert Advisory Group. Expert advisory group report 3: Toward a world-class health data system (2022). Accessed 7 June 2025.

[27] Heo, S. *et al.* Impacts of the covid-19 pandemic on scientists' productivity in science, technology, engineering, mathematics (stem), and medicine fields. *Humanities and Social Sciences Communications* **9**, 1–11 (2022).

[28] Office of the United States Trade Representative. United states–mexico–canada agreement (usmca), chapter 19: Digital trade (2020). URL https://ustr.gov/sites/default/files/files/agreements/FTA/USMCA/Text/19-Digital-Trade.pdf. Articles 19.11 (Cross-Border Transfer of Information) and 19.12 (Location of Computing Facilities).

[29] G7 international patient summary roadmap. Tech. Rep., UK Department of Health and Social Care (2021). URL https://assets.publishing.service.gov.uk/media/61d82fbd8fa8f505893f1c93/G7-international-patient-summary-roadmap.pdf.

[30] Facilitating the Secondary Use of Health Data for Public Interest Purposes Across Borders. Tech. Rep., Organisation for Economic Co-operation and Development (2025).

[31] Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626 (2017).

[32] Zhang, H. & Qie, Y. Applying deep learning to medical imaging: a review. *Applied Sciences* **13**, 10521 (2023).

[33] Ibrahim, M. *et al.* Generative ai for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges. *arXiv preprint arXiv:2407.00116* (2024).

[34] Kush, R. D. *et al.* Fair data sharing: the roles of common data elements and harmonization. *Journal of biomedical informatics* **107**, 103421 (2020).

[35] Friesen, P. *et al.* Governing ai-driven health research: Are irbs up to the task? *Ethics & Human Research* **43**, 35–42 (2021).

[36] Althnian, A. *et al.* Impact of dataset size on classification performance: an empirical evaluation in the medical domain. *Applied Sciences* **11**, 796 (2021).

[37] Decherchi, S., Pedrini, E., Mordenti, M., Cavalli, A. & Sangiorgi, L. Opportunities and challenges for machine learning in rare diseases. *Frontiers in medicine* **8**, 747612 (2021).

[38] Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): the tripod statement. *Journal of British Surgery* **102**, 148–158 (2015).

[39] Christodoulou, E. *et al.* A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology* **110**, 12–22 (2019).

[40] Canada Health Infoway. Shared pan-canadian interoperability roadmap. Tech. Rep., Canada Health Infoway (2023).

[41] Turnbull, C. *et al.* The 100 000 genomes project: bringing whole genome sequencing to the nhs. *Bmj* **361** (2018).

[42] Sosinsky, A. *et al.* Insights for precision oncology from the integration of genomic and clinical data of 13,880 tumors from the 100,000 genomes cancer programme. *Nature medicine* **30**, 279–289 (2024).

[43] 100, . G. P. P. I. 100,000 genomes pilot on rare-disease diagnosis in health care—preliminary report. *New England Journal of Medicine* **385**, 1868–1880 (2021).

[44] Mayo, K. R. *et al.* The all of us data and research center: creating a secure, scalable, and sustainable ecosystem for biomedical research. *Annual review of biomedical data science* **6**, 443–464 (2023).

[45] Sudlow, C. *et al.* Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12**, e1001779 (2015).

[46] Singhal, K. *et al.* Large language models encode clinical knowledge. *Nature* **620**, 172–180 (2023).

[47] Yang, X. *et al.* Gatortron: A large clinical language model to unlock patient information from unstructured electronic health records. *arXiv preprint arXiv:2203.03540* (2022).

[48] Ahmed, F. *et al.* Pathalign: A vision-language model for whole slide images in histopathology. *arXiv preprint arXiv:2406.19578* (2024).

[49] Pellegrini, C. *et al.* Radialog: Large vision-language models for x-ray reporting and dialog-driven assistance. In *Medical Imaging with Deep Learning* (2025).

[50] Carlini, N. *et al.* Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)*, 2633–2650 (USENIX Association, 2021). URL https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting.

[51] Murphy, S. N. *et al.* Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *Journal of the American Medical Informatics Association* **17**, 124–130 (2010).

[52] Verma, A. A. *et al.* Patient characteristics, resource use and outcomes associated with general internal medicine hospital care: the general medicine inpatient initiative (gemini) retrospective cohort study. *Canadian Medical Association Open Access Journal* **5**, E842–E849 (2017).

[53] Sweeney, L. Only you, your doctor, and many others may know. *Technology Science* **2015092903**, 29 (2015).

[54] Dwork, C., Roth, A. *et al.* The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* **9**, 211–407 (2014).

[55] Goncalves, A. *et al.* Generation and evaluation of synthetic patient data. *BMC medical research methodology* **20**, 1–40 (2020).

[56] Gentry, C. Fully homomorphic encryption using ideal lattices. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, 169–178 (2009).

[57] Cramer, R., Damgård, I. B. *et al. Secure multiparty computation* (Cambridge University Press, 2015).

[58] Qi, Y. *et al.* Simulating federated learning for steatosis detection using ultrasound images. *Scientific Reports* **14**, 13253 (2024).

[59] Lhatoo, S. D. *et al.* Big data in epilepsy: clinical and research considerations. report from the epilepsy big data task force of the international league against epilepsy. *Epilepsia* **61**, 1869–1883 (2020).

[60] Roth, H. R. *et al.* Nvidia flare: Federated learning from simulation to real-world. *arXiv preprint arXiv:2210.13291* (2022).

[61] Chomutare, T. *et al.* Artificial intelligence implementation in healthcare: a theory-based scoping review of barriers and facilitators. *International Journal of Environmental Research and Public Health* **19**, 16359 (2022).

[62] Nair, M., Svedberg, P., Larsson, I. & Nygren, J. M. A comprehensive overview of barriers and strategies for ai implementation in healthcare: Mixed-method design. *Plos one* **19**, e0305949 (2024).

[63] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G. & King, D. Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine* **17**, 1–9 (2019).

[64] National Library of Medicine. List of all journals cited in PubMed (2021). URL https://www.nlm.nih.gov/bsd/serfile_addedinfo.html. Last reviewed: June 10, 2021.

[65] Kans, J. Entrez direct: E-utilities on the unix command line. In *Entrez programming utilities help [Internet]* (National Center for Biotechnology Information (US), 2024). URL https://www.ncbi.nlm.nih.gov/books/NBK179288/.

[66] U.S. National Library of Medicine. Medline pubmed xml element descriptions and their attributes. URL https://www.nlm.nih.gov/bsd/licensee/elements_descriptions.html. Accessed: 2024-09-23.

[67] Touvron, H. *et al.* Llama: Open and efficient foundation language models (2023). URL https://arxiv.org/abs/2302.13971. 2302.13971.

[68] Chase, H. Langchain: Building applications with llms through composability (2022). URL https://github.com/langchain-ai/langchain.

[69] Johnson, A. E. *et al.* Mimic-iv, a freely accessible electronic health record dataset. *Scientific Data* **10** (2023).

[70] Lewis, P. *et al.* Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* **33**, 9459–9474 (2020).

[71] Liu, H., Li, C., Wu, Q. & Lee, Y. J. Visual instruction tuning. In *NeurIPS* (2023).

Supplementary Information

Precision Medicine in Canada: A Scoping Review of AI Collaboration across Patient Level and Multimodal Data

Omid Jafarinezhad, Qian Zhang, Ryan Rezai, Mohammad Noaeen, Aviv Shachak, Behrouz Far, Zahra Shakeri

Contents

1 Supplementary Note 1: Search Strategy Details

A targeted search strategy was designed to identify relevant studies from the PubMed database using the following query:

> (machine learning **OR** artificial intelligence) **NOT** (review[PTYP] **OR** editorial[PTYP] **OR** systematic review[PTYP] **OR** retracted publication[PTYP] **OR** retraction of publication[PTYP]) **AND** 2018/01/01:2025/02/28[PDAT]

This query was structured to retrieve studies containing 'machine learning' or 'artificial intelligence' across all fields in PubMed, ensuring comprehensive coverage of AI-driven healthcare research. Non-original research articles, such as reviews, editorials, systematic reviews, and retracted publications, were excluded using the 'NOT' operator and the 'PTYP' (publication type) filter. A date filter (2018/01/01 to 2025/02/28) was applied to limit the search to recent publications, enabling analysis of contemporary trends in AI-driven healthcare research. The search was executed using Entrez Direct (EDirect) for querying PubMed, and XML records were retrieved via E-Utilities. Key metadata, including PMID, title, abstract, keywords, MeSH terms, authors, affiliations, journal, and publication year, were extracted for further analysis.

1.1 Identification of Canadian Collaborators

Each paper contained a nested structure of free-text affiliation strings, as multiple authors could have several affiliations. These affiliation strings were used to determine whether Canadian institutional affiliations were present. If Canadian affiliations were identified, the paper was then categorized as either 'domestic collaboration' or 'international collaboration' (as defined in Section). However, the classification process was complicated by variations in the affiliation texts, which made accurate detection difficult. For instance, affiliations such as 'Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada' clearly indicated a Canadian affiliation and were straightforward to identify. However, other affiliations, such as 'School of Continuous Studies, McGill University' or 'Department

of Health Sciences Research, Rochester MN,' presented different formats, abbreviations, or incomplete information, which required more detailed processing to ascertain whether the affiliation was Canadian.

In the first stage of the filtering, affiliation texts with clear geographic information were handled using exact word matching for terms such as 'Canada' or 'Toronto'. This method efficiently identified papers with at least one Canadian-affiliated author. For papers with more ambiguous or incomplete geographic information, the LLaMA 3 language model was employed to analyze the affiliation texts and identify more subtle indicators of Canadian institutions. The model processed the affiliation list for each paper and returned a list of countries associated with each affiliation, thereby facilitating more accurate identification of Canadian affiliations.

2 Supplementary Note 2: Technical Pipeline

2.1 Detection of the patient-level data usage

The pipeline operated using a structured prompt to extract specific information from each paper. The prompt provided to the model was as follows:

> Given the following paper details:
> - **Title**: {Title}
> - **Abstract**: {Abstract}
> - **Affiliations**: {Affiliations}
>
> Return a JSON object with the following fields:
> - **affiliation_countries**: A List of countries associated with the institutions in the affiliations.
> - **is_patient_level_data**: A boolean indicating if the paper works with patient-level data.
> - **patient_data_type**: The type of patient-level data used (e.g., genomic, clinical, imaging).
>
> Return only the JSON object without any additional text.

The LLaMA3-based pipeline returned results in the specified JSON format, allowing for the systematic identification of papers that involved patient-level data, as well as extracting key information about the institutions and countries involved in the data collection process.

2.2 Feature Extraction pipeline

To ensure comprehensive analysis, we retrieved the full text of each paper in the dataset, which enabled the extraction of detailed information such as data collection methods. The full texts, along with accompanying figures and tables, were then converted into vector representations using a pre-trained embedding model. These vector embeddings were stored in a vector database, allowing for efficient context-aware retrieval during the extraction process. This approach provided both scalability and precision, ensuring rapid access to relevant information while maintaining accuracy in handling large volumes of unstructured data.

With the vector database in place, we employed a RAG model, utilizing LLMs to extract collaboration features in a structured JSON format. Carefully designed prompts guided the LLM to retrieve and organize relevant information from the vector database, transforming complex, unstructured data from full-text documents into a machine-readable format. This process ensured consistency and precision in feature extraction.

To validate the accuracy of the extracted features, we ran the pipeline using both LLaMA 3 and LLaMA 3.1 (7B parameters) and conducted a manual review on a random sample of 750 papers. The validation involved comparing the LLM-generated outputs with human-extracted data to ensure the reliability of the structured feature extraction pipeline. Based on the validation results, we applied a

maturity voting mechanism to identify any discrepancies and further refine the pipeline. This iterative process optimized the accuracy of the pipeline, balancing both automated extraction and human oversight to ensure reliable outputs.

2.3 Data Cleaning and Preprocessing

After the feature extraction process, the collected data and features were transferred to a PostgreSQL database for comprehensive cleaning and preprocessing. This stage was crucial to ensure the reliability, consistency, and quality of the data for further analysis. Key tasks during this stage included handling incomplete data, merging datasets, and normalizing feature values to create a robust dataset for analysis.

We employed multiple prompts within our pipeline to extract different features from the dataset. Each prompt targeted specific aspects, such as patient-level data, data modalities, collaboration countries, and data-sharing practices. This approach allowed us to systematically retrieve key features relevant to our analysis. Since different features were extracted from separate parts of the data, the resulting information came from multiple datasets. To create a unified dataset, we merged these datasets based on publication IDs. By merging these datasets, we ensured that all relevant information was aggregated into a single, comprehensive dataset.

Given that the data extraction was conducted using LLMs, there were instances where certain features were not explicitly detailed in the papers. In such cases, the model produced responses like 'Not specified'. During preprocessing, these placeholders were mapped to NULL values to ensure that incomplete or non-informative data did not skew our results.

Additionally, for fields such as collaboration countries or data modalities, we applied a normalization process to convert any array-like structures into sets of unique values. This eliminated duplicates, ensuring that entries, such as repeated country names or modalities, were counted only once. By doing so, we preserved the integrity of the dataset and avoided potential over-counting of specific features.

3 Supplementary References: Included Articles

This section provides a complete list of all studies included in the review.[a1–a160] The numbering corresponds directly to Table 1 as well as to the references cited in the main text.

[a1] Namavarian, A. *et al.* Length of stay prediction models for oral cancer surgery: Machine learning, statistical and acs-nsqip. *The Laryngoscope* **134**, 3664–3672 (2024).

[a2] Kim, B.-G. *et al.* White matter diffusion estimates in obsessive-compulsive disorder across 1653 individuals: machine learning findings from the enigma ocd working group. *Molecular Psychiatry* **29**, 1063–1074 (2024).

[a3] Kushol, R. *et al.* Sf2former: Amyotrophic lateral sclerosis identification from multi-center mri data using spatial and frequency fusion transformer. *Computerized Medical Imaging and Graphics* **108**, 102279 (2023).

[a4] Petch, J. *et al.* Optimizing warfarin dosing for patients with atrial fibrillation using machine learning. *Scientific reports* **14**, 4516 (2024).

[a5] Nunes, A. *et al.* Using structural mri to identify bipolar disorders–13 site machine learning study in 3020 individuals from the enigma bipolar disorders working group. *Molecular psychiatry* **25**, 2130–2143 (2020).

[a6] Schapranow, M.-P. *et al.* Nephrocage—german-canadian consortium on ai for improved kidney transplantation outcome: Protocol for an algorithm development and validation study. *JMIR Research Protocols* **12**, e48892 (2023).

[a7] Pati, S. *et al.* Federated learning enables big data for rare cancer boundary detection. *Nature communications* **13**, 7346 (2022).

[a8] Warkentin, M. T. *et al.* Radiomics analysis to predict pulmonary nodule malignancy using machine learning approaches. *Thorax* **79**, 307–315 (2024).

[a9] Halloran, K. *et al.* Molecular t-cell–mediated rejection in transbronchial and mucosal lung transplant biopsies is associated with future risk of graft loss. *The Journal of Heart and Lung Transplantation* **39**, 1327–1337 (2020).

[a10] Mallon, A.-M. *et al.* Advancing data science in drug development through an innovative computational framework for data sharing and statistical analysis. *BMC medical research methodology* **21**, 250 (2021).

[a11] Salle, F. G. *et al.* Comprehensive molecular and pathologic evaluation of transitional mesothelioma assisted by deep learning approach: a multi-institutional study of the international mesothelioma panel from the mesopath reference center. *Journal of thoracic oncology* **15**, 1037–1053 (2020).

[a12] Bogowicz, M. *et al.* Privacy-preserving distributed learning of radiomics to predict overall survival and hpv status in head and neck cancer. *Scientific reports* **10**, 4542 (2020).

[a13] Jibb, L. A. *et al.* Traumatic stress symptoms in family caregivers of patients with acute leukaemia: protocol for a multisite mixed methods, longitudinal, observational study. *BMJ open* **12**, e065422 (2022).

[a14] Mahmoodifar, S. *et al.* Comparative analysis of the spatial distribution of brain metastases across several primary cancers using machine learning and deep learning models. *Journal of Neuro-Oncology* **167**, 501–508 (2024).

[a15] Iv, M. *et al.* Mr imaging–based radiomic signatures of distinct molecular subgroups of medulloblastoma. *American Journal of Neuroradiology* **40**, 154–161 (2019).

[a16] Emaus, M. J. *et al.* Bragatston study protocol: a multicentre cohort study on automated quantification of cardiovascular calcifications on radiotherapy planning ct scans for cardiovascular risk prediction in patients with breast cancer. *BMJ open* **9**, e028752 (2019).

[a17] Tsilimigras, D. I. *et al.* A machine-based approach to preoperatively identify patients with the most and least benefit associated with resection for intrahepatic cholangiocarcinoma: an international multi-institutional analysis of 1146 patients. *Annals of surgical oncology* **27**, 1110–1119 (2020).

[a18] Iantsen, A. *et al.* Convolutional neural networks for pet functional volume fully automatic segmentation: development and validation in a multi-center setting. *European journal of nuclear medicine and molecular imaging* **48**, 3444–3456 (2021).

[a19] Quon, J. *et al.* Deep learning for pediatric posterior fossa tumor detection and classification: a multi-institutional study. *American Journal of Neuroradiology* **41**, 1718–1725 (2020).

[a20] Yolchuyeva, S. *et al.* A radiomics-clinical model predicts overall survival of non-small cell lung cancer patients treated with immunotherapy: a multicenter study. *Cancers* **15**, 3829 (2023).

[a21] van Dijk, L. V. *et al.* Head and neck cancer predictive risk estimator to determine control and therapeutic outcomes of radiotherapy (hnc-predictor): development, international multi-institutional validation, and web implementation of clinic-ready model-based risk stratification for head and neck cancer. *European Journal of Cancer* **178**, 150–161 (2023).

[a22] Bratulic, S. *et al.* Noninvasive detection of any-stage cancer using free glycosaminoglycans. *Proceedings of the National Academy of Sciences* **119**, e2115328119 (2022).

[a23] Randall, J. *et al.* Image-based deep neural network for individualizing radiotherapy dose is transportable across health systems. *JCO Clinical Cancer Informatics* **7**, e2200100 (2023).

[a24] Dafaalla, M. *et al.* Outcomes of st elevation myocardial infarction in patients with cancer: a nationwide study. *European Heart Journal-Quality of Care and Clinical Outcomes* **9**, 806–817 (2023).

[a25] Petch, J. *et al.* Developing a data and analytics platform to enable a breast cancer learning health system at a regional cancer center. *JCO Clinical Cancer Informatics* **7**, e2200182 (2023).

[a26] DeVries, D. A. *et al.* Dual-center validation of using magnetic resonance imaging radiomics to predict stereotactic radiosurgery outcomes. *Neuro-Oncology Advances* **5**, vdad064 (2023).

[a27] Kazmierski, M. *et al.* Multi-institutional prognostic modeling in head and neck cancer: evaluating impact and generalizability of deep learning and radiomics. *Cancer Research Communications* **3**, 1140–1151 (2023).

[a28] Yolchuyeva, S. *et al.* Radiomics approaches to predict pd-l1 and pfs in advanced non-small cell lung patients treated with immunotherapy: a multi-institutional study. *Scientific reports* **13**, 11065 (2023).

[a29] Tonneau, M. *et al.* Generalization optimizing machine learning to improve ct scan radiomics and assess immune checkpoint inhibitors' response in non-small cell lung cancer: a multicenter cohort study. *Frontiers in Oncology* **13**, 1196414 (2023).

[a30] Soldatelli, M. D. *et al.* Identification of multiclass pediatric low-grade neuroepithelial tumor molecular subtype with adc mr imaging and machine learning. *American Journal of Neuroradiology* **45**, 753–760 (2024).

[a31] Yolchuyeva, S. *et al.* Multi-institutional prognostic modeling of survival outcomes in nsclc patients treated with first-line immunotherapy using radiomics. *Journal of translational medicine* **22**, 42 (2024).

[a32] Ahmed, S. R. *et al.* Reproducible and clinically translatable deep neural networks for cervical screening. *Scientific reports* **13**, 21772 (2023).

[a33] Dawson, H. *et al.* Lymph node metastases and recurrence in pt1 colorectal cancer: Prediction with the international budding consortium score—a retrospective, multi-centric study. *United European Gastroenterology Journal* **12**, 299–308 (2024).

[a34] Taha-Mehlitz, S. *et al.* Machine learning-based preoperative analytics for the prediction of anastomotic leakage in colorectal surgery: a swiss pilot study. *Surgical Endoscopy* **38**, 3672–3683 (2024).

[a35] Wilson, P. F. *et al.* Toward confident prostate cancer detection using ultrasound: a multi-center study. *International Journal of Computer Assisted Radiology and Surgery* **19**, 841–849 (2024).

[a36] Hyndman, M. E. *et al.* Development of an effective predictive screening tool for prostate cancer using the claritydx machine learning platform. *NPJ Digital Medicine* **7**, 163 (2024).

[a37] Asadi-Aghbolaghi, M. *et al.* Learning generalizable ai models for multi-center histopathology image classification. *NPJ Precision Oncology* **8**, 151 (2024).

[a38] Altaf, A. *et al.* Development of an artificial intelligence–based model to predict early recurrence of neuroendocrine liver metastasis after resection. *Journal of Gastrointestinal Surgery* **28**, 1828–1837 (2024).

[a39] Tschandl, P., Argenziano, G., Razmara, M. & Yap, J. Diagnostic accuracy of content-based dermatoscopic image retrieval with deep classification features. *British Journal of Dermatology* **181**, 155–165 (2019).

[a40] Li, B. *et al.* Predicting outcomes following lower extremity open revascularization using machine learning. *Scientific Reports* **14**, 2899 (2024).

[a41] Betancur, J. *et al.* Deep learning analysis of upright-supine high-efficiency spect myocardial perfusion imaging for prediction of obstructive coronary artery disease: a multicenter study. *Journal of Nuclear Medicine* **60**, 664–670 (2019).

[a42] Dundas, J. *et al.* Interaction of ai-enabled quantitative coronary plaque volumes on coronary ct angiography, ffrct, and clinical outcomes: a retrospective analysis of the advance registry. *Circulation: Cardiovascular Imaging* **17**, e016143 (2024).

[a43] Betancur, J. *et al.* Deep learning for prediction of obstructive disease from fast myocardial perfusion spect: a multicenter study. *JACC: Cardiovascular Imaging* **11**, 1654–1663 (2018).

[a44] Xu, C. *et al.* Segmentation and quantification of infarction without contrast agents via spatiotemporal generative adversarial learning. *Medical image analysis* **59**, 101568 (2020).

[a45] Xu, Y. *et al.* Enhancing icd-code-based case definition for heart failure using electronic medical record data. *Journal of Cardiac Failure* **26**, 610–617 (2020).

[a46] Tromp, J. *et al.* Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study. *The Lancet Digital Health* **4**, e46–e54 (2022).

[a47] Miller, R. J. *et al.* Ai-defined cardiac anatomy improves risk stratification of hybrid perfusion imaging. *Cardiovascular Imaging* **17**, 780–791 (2024).

[a48] Eisenberg, E. *et al.* Diagnostic safety of a machine learning-based automatic patient selection algorithm for stress-only myocardial perfusion spect. *Journal of Nuclear Cardiology* **29**, 2295–2307 (2022).

[a49] Boyko, M. *et al.* Prevalence of high-risk aortic arch atherosclerosis features on computed tomography angiography in embolic stroke of undetermined source. *Journal of Stroke and Cerebrovascular Diseases* **32**, 107374 (2023).

[a50] Brown, J. R. *et al.* Information extraction from electronic health records to predict readmission following acute myocardial infarction: does natural language processing using clinical notes improve prediction of readmission? *Journal of the American Heart Association* **11**, e024198 (2022).

[a51] Rios, R. *et al.* Handling missing values in machine learning to predict patient-specific risk of adverse cardiac events: Insights from refine spect registry. *Computers in biology and medicine* **145**, 105449 (2022).

[a52] Tamarappoo, B. K. *et al.* Differences in prognostic value of myocardial perfusion single-photon emission computed tomography using high-efficiency solid-state detector between men and women in a large international multicenter study. *Circulation: Cardiovascular Imaging* **15**, e012741 (2022).

[a53] Bhavnani, S. P. *et al.* Multicenter validation of a machine learning phase space electro-mechanical pulse wave analysis to predict elevated left ventricular end diastolic pressure at the point-of-care. *PLoS One* **17**, e0277300 (2022).

[a54] Wu, Z. *et al.* Prediction of preoperative in-hospital mortality rate in patients with acute aortic dissection by machine learning: a two-centre, retrospective cohort study. *BMJ open* **13**, e066782 (2023).

[a55] Tzimas, G. *et al.* Age-and sex-specific nomographic ct quantitative plaque data from a large international cohort. *Cardiovascular Imaging* **17**, 165–175 (2024).

[a56] Mor-Avi, V. *et al.* Deep learning assisted measurement of echocardiographic left heart parameters: improvement in interobserver variability and workflow efficiency. *The International Journal of Cardiovascular Imaging* **39**, 2507–2516 (2023).

[a57] Nedadur, R. *et al.* Machine learning and decision making in aortic arch repair. *The Journal of Thoracic and Cardiovascular Surgery* **169**, 59–67 (2025).

[a58] Miller, R. J. *et al.* Simplified approach to predicting obstructive coronary disease with integration of coronary calcium: development and external validation. *Journal of the American Heart Association* **12**, e031601 (2023).

[a59] Mansouri, M. *et al.* Ct analysis of aortic calcifications to predict abdominal aortic aneurysm rupture. *European Radiology* **34**, 3903–3911 (2024).

[a60] Fard, S. S., Perkins, T. J. & Wells, P. S. Machine learning analysis of bleeding status in venous thromboembolism patients. *Research and Practice in Thrombosis and Haemostasis* **8**, 102403 (2024).

[a61] Domingo-Gardeta, T. *et al.* Rationale and design of the artificial intelligence scalable solution for acute myocardial infarction (assist) study. *Journal of Electrocardiology* **86**, 153768 (2024).

[a62] Mayourian, J. *et al.* Deep learning-based electrocardiogram analysis predicts biventricular dysfunction and dilation in congenital heart disease. *Journal of the American College of Cardiology* **84**, 815–828 (2024).

[a63] Yi, J. *et al.* Ai-based volumetric six-tissue body composition quantification from ct cardiac attenuation scans for mortality prediction: a multicentre study. *The Lancet Digital Health* **7** (2025).

[a64] Agrawal, V. *et al.* Federated learning and differential privacy techniques on multi-hospital population-scale electrocardiogram data. In *Proceedings of the 2024 8th International Conference on Medical and Health Informatics*, 143–152 (2024).

[a65] Arroyo-Gallego, T. *et al.* Detecting motor impairment in early parkinson's disease via natural typing interaction with keyboards: Validation of the neuroqwerty approach in an uncontrolled at-home setting. *Journal of medical Internet research* **20**, e89 (2018).

[a66] Trebaul, L. *et al.* Probabilistic functional tractography of the human cortex revisited. *Neuroimage* **181**, 414–429 (2018).

[a67] Merali, Z. G., Witiw, C. D., Badhiwala, J. H., Wilson, J. R. & Fehlings, M. G. Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy. *PloS one* **14**, e0215133 (2019).

[a68] Mithani, K. *et al.* Connectomic profiling identifies responders to vagus nerve stimulation. *Annals of neurology* **86**, 743–753 (2019).

[a69] Khademi, A., Reiche, B., DiGregorio, J., Arezza, G. & Moody, A. R. Whole volume brain extraction for multi-centre, multi-disease flair mri datasets. *Magnetic Resonance Imaging* **66**, 116–130 (2020).

[a70] Warsi, N. M. *et al.* Dissociable default-mode subnetworks subserve childhood attention and cognitive flexibility: Evidence from deep learning and stereotactic electroencephalography. *Neural Networks* **167**, 827–837 (2023).

[a71] Zhang, M. *et al.* Machine-learning approach to differentiation of benign and malignant peripheral nerve sheath tumors: a multicenter study. *Neurosurgery* **89**, 509–517 (2021).

[a72] Mouches, P., Wilms, M., Rajashekar, D., Langner, S. & Forkert, N. D. Multimodal biological brain age prediction using magnetic resonance imaging and angiography with the identification of predictive regions. *Human brain mapping* **43**, 2554–2566 (2022).

[a73] Basiratzadeh, S. *et al.* A data-driven approach to categorize patients with traumatic spinal cord injury: cluster analysis of a multicentre database. *Frontiers in Neurology* **14**, 1263291 (2023).

[a74] Kernbach, J. M. *et al.* Bayesian stroke modeling details sex biases in the white matter substrates of aphasia. *Communications biology* **6**, 354 (2023).

[a75] Tröger, J. *et al.* Prosa—a multicenter prospective observational study to develop low-burden digital speech biomarkers in als and ftd. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* **24**, 589–598 (2023).

[a76] Theaud, G. *et al.* Doris: A diffusion mri-based 10 tissue class deep learning segmentation algorithm tailored to improve anatomically-constrained tractography. *Frontiers in Neuroimaging* **1**, 917806 (2022).

[a77] Totzek, J. F. *et al.* Longitudinal inference of multiscale markers in psychosis: from hippocampal centrality to functional outcome. *Molecular psychiatry* **29**, 2929–2938 (2024).

[a78] Sen, B., Borle, N. C., Greiner, R. & Brown, M. R. A general prediction model for the detection of adhd and autism using structural and functional mri. *PloS one* **13**, e0194856 (2018).

[a79] Bohn, L. *et al.* Machine learning analyses identify multi-modal frailty factors that selectively discriminate four cohorts in the alzheimer's disease spectrum: a compass-nd study. *BMC geriatrics* **23**, 837 (2023).

[a80] Wang, H. *et al.* Whole-genome sequencing analysis reveals new susceptibility loci and structural variants associated with progressive supranuclear palsy. *Molecular neurodegeneration* **19**, 61 (2024).

[a81] Souza, R. *et al.* A multi-center distributed learning approach for parkinson's disease classification using the traveling model paradigm. *Frontiers in Artificial Intelligence* **7**, 1301997 (2024).

[a82] Zabihi, M. *et al.* Nonlinear latent representations of high-dimensional task-fmri data: Unveiling cognitive and behavioral insights in heterogeneous spatial maps. *Plos one* **19**, e0308329 (2024).

[a83] Olafson, E. R. *et al.* Data-driven biomarkers better associate with stroke motor outcomes than theory-based biomarkers. *Brain communications* **6**, fcae254 (2024).

[a84] Park, J. *et al.* Gene–environment pathways to cognitive intelligence and psychotic-like experiences in children. *Elife* **12**, RP88117 (2024).

[a85] Afzali, M. H. *et al.* Machine-learning prediction of adolescent alcohol use: A cross-study, cross-cultural validation. *Addiction* **114**, 662–671 (2019).

[a86] Toenders, Y. J. *et al.* Predicting depression onset in young people based on clinical, cognitive, environmental, and neurobiological data. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **7**, 376–384 (2022).

[a87] Group, E. O. W. *et al.* Structural neuroimaging biomarkers for obsessive-compulsive disorder in the enigma-ocd consortium: medication matters. *Translational psychiatry* **10**, 342 (2020).

[a88] Ballester, P. L. *et al.* Accelerated brain aging in major depressive disorder and antidepressant treatment response: A can-bind report. *NeuroImage: Clinical* **32**, 102864 (2021).

[a89] Cearns, M. *et al.* Using polygenic scores and clinical data for bipolar disorder patient stratification and lithium response prediction: machine learning approach. *The British Journal of Psychiatry* **220**, 219–228 (2022).

[a90] Benrimoh, D. *et al.* Towards outcome-driven patient subgroups: a machine learning analysis across six depression treatment studies. *The American Journal of Geriatric Psychiatry* **32**, 280–292 (2024).

[a91] Belov, V. *et al.* Multi-site benchmark classification of major depressive disorder using machine learning on cortical and subcortical measures. *Scientific reports* **14**, 1084 (2024).

[a92] Nogovitsyn, N. *et al.* An empirical analysis of structural neuroimaging profiles in a staging model of depression. *Journal of Affective Disorders* **351**, 631–640 (2024).

[a93] Zhu, Y. *et al.* Using brain structural neuroimaging measures to predict psychosis onset for individuals at clinical high-risk. *Molecular psychiatry* **29**, 1465–1477 (2024).

[a94] Dong, M. S. *et al.* Multimodal workflows optimally predict response to repetitive transcranial magnetic stimulation in patients with schizophrenia: a multisite machine learning analysis. *Translational Psychiatry* **14**, 196 (2024).

[a95] Ho, N. C. *et al.* Atypical brain aging and its association with working memory performance in major depressive disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **9**, 786–799 (2024).

[a96] Perlman, K. *et al.* Development of a differential treatment selection model for depression on consolidated and transformed clinical trial datasets. *Translational Psychiatry* **14**, 263 (2024).

[a97] Joel, S. *et al.* Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 19061–19071 (2020).

[a98] N. Fountoulakis, K. *et al.* Results of the covid-19 mental health international for the health professionals (comet-hp) study: depression, suicidal tendencies and conspiracism. *Social psychiatry and psychiatric epidemiology* **58**, 1387–1410 (2023).

[a99] Murphy, C. E. *et al.* A targeted strategic peer support intervention to increase adherence to video teletherapy exposure and response prevention treatment for obsessive-compulsive disorder: a retrospective observational analysis. *Frontiers in Human Neuroscience* **17**, 1251194 (2023).

[a100] Vince, N. *et al.* Association of hla-drb1∗ 09: 01 with tige levels among african-ancestry individuals with asthma. *Journal of Allergy and Clinical Immunology* **146**, 147–155 (2020).

[a101] Zafari, H., Langlois, S., Zulkernine, F., Kosowan, L. & Singer, A. Ai in predicting copd in the canadian population. *Biosystems* **211**, 104585 (2022).

[a102] Arntfield, R. *et al.* Automation of lung ultrasound interpretation via deep learning for the classification of normal versus abnormal lung parenchyma: a multicenter study. *Diagnostics* **11**, 2049 (2021).

[a103] Li, D., Lin, C. T., Sulam, J. & Yi, P. H. Deep learning prediction of sex on chest radiographs: a potential contributor to biased algorithms. *Emergency Radiology* **29**, 365–370 (2022).

[a104] He, P. *et al.* Early prediction of pediatric asthma in the canadian healthy infant longitudinal development (child) birth cohort using machine learning. *Pediatric Research* **95**, 1818–1825 (2024).

[a105] Sterling, K. L. *et al.* Healthcare resource utilisation and costs in patients with treated obstructive sleep apnea. *Journal of Sleep Research* **33**, e14099 (2024).

[a106] Angelini, E. D. *et al.* Pulmonary emphysema subtypes defined by unsupervised machine learning on ct scans. *Thorax* **78**, 1067–1079 (2023).

[a107] Fakhraei, R. *et al.* Development of multivariable prediction models for the identification of patients admitted to hospital with an exacerbation of copd and the prediction of risk of readmission: a retrospective cohort study using electronic medical record data. *COPD: Journal of Chronic Obstructive Pulmonary Disease* **20**, 274–283 (2023).

[a108] Wu, D. *et al.* Improving the generalizability and performance of an ultrasound deep learning model using limited multicenter data for lung sliding artifact identification. *Diagnostics* **14**, 1081 (2024).

[a109] Hartl, W. H. *et al.* Relevance of protein intake for weaning in the mechanically ventilated critically ill: analysis of a large international database. *Critical Care Medicine* **52**, e121–e131 (2024).

[a110] Semmad, A. & Bahoura, M. Comparative study of respiratory sounds classification methods based on cepstral analysis and artificial neural networks. *Computers in Biology and Medicine* **171**, 108190 (2024).

[a111] Abdulahi, A. T., Ogundokun, R. O., Adenike, A. R., Shah, M. A. & Ahmed, Y. K. Pulmonet: a novel deep learning based pulmonary diseases detection model. *BMC Medical Imaging* **24**, 51 (2024).

[a112] Moran-Mendoza, O., Singla, A., Kalra, A., Muelly, M. & Reicher, J. J. Computed tomography machine learning classifier correlates with mortality in interstitial lung disease. *Respiratory Investigation* **62**, 670–676 (2024).

[a113] Liu, P. *et al.* Predicting the risks of kidney failure and death in adults with moderate to severe chronic kidney disease: multinational, longitudinal, population based, cohort study. *bmj* **385** (2024).

[a114] Ferguson, T. *et al.* Development and external validation of a machine learning model for progression of ckd. *Kidney International Reports* **7**, 1772–1781 (2022).

[a115] Mosa, A. I., Watts, D. & Tangri, N. Impacting management of chronic kidney disease through primary care practice audits: a quality improvement study. *Canadian Journal of Kidney Health and Disease* **9**, 20543581221144840 (2022).

[a116] Aubert, O. *et al.* Archetype analysis identifies distinct profiles in renal transplant recipients with transplant glomerulopathy associated with allograft survival. *Journal of the American Society of Nephrology* **30**, 625–639 (2019).

[a117] Abdullah, S. S. *et al.* High-throughput computing to automate population-based studies to detect the 30-day risk of adverse outcomes after new outpatient medication use in older adults with chronic kidney disease: A clinical research protocol. *Canadian Journal of Kidney Health and Disease* **11**, 20543581231221891 (2024).

[a118] Tangri, N. *et al.* Machine learning for prediction of chronic kidney disease progression: validation of the klinrisk model in the canvas program and credence trial. *Diabetes, Obesity and Metabolism* **26**, 3371–3380 (2024).

[a119] Dayan, I. *et al.* Federated learning for predicting clinical outcomes in patients with covid-19. *Nature medicine* **27**, 1735–1743 (2021).

[a120] Gourdeau, D. *et al.* Tracking and predicting covid-19 radiological trajectory on chest x-rays using deep learning. *Scientific reports* **12**, 5616 (2022).

[a121] Ferreira, V. H. *et al.* Omicron ba. 4/5 neutralization and t-cell responses in organ transplant recipients after booster messenger rna vaccine: a multicenter cohort study. *Clinical Infectious Diseases* **77**, 229–236 (2023).

[a122] de Andrade Rodrigues, R. S. *et al.* Prediction of the occurrence of leprosy reactions based on bayesian networks. *Frontiers in Medicine* **10**, 1233220 (2023).

[a123] Bai, A. D. *et al.* Comparative effectiveness of first-line and alternative antibiotic regimens in hospitalized patients with nonsevere community-acquired pneumonia: a multicenter retrospective cohort study. *Chest* **165**, 68–78 (2024).

[a124] McBurney, S. H. *et al.* Validating pertussis data measures using electronic medical record data in ontario, canada 1986–2016. *Vaccine: X* **15**, 100408 (2023).

[a125] Villar, J. *et al.* Predicting the length of mechanical ventilation in acute respiratory disease syndrome using machine learning: the pioneer study. *Journal of Clinical Medicine* **13**, 1811 (2024).

[a126] Mamandipoor, B. *et al.* Machine learning predicts mortality based on analysis of ventilation parameters of critically ill patients: multi-centre validation. *BMC Medical Informatics and Decision Making* **21**, 152 (2021).

[a127] Wang, L. *et al.* Boosting delirium identification accuracy with sentiment-based natural language processing: mixed methods study. *JMIR Medical Informatics* **10**, e38161 (2022).

[a128] Villar, J. *et al.* Predicting icu mortality in acute respiratory distress syndrome patients using machine learning: the predicting outcome and stratification of severity in ards (postcards) study. *Critical care medicine* **51**, 1638–1649 (2023).

[a129] Mutnuri, M. K., Stelfox, H. T., Forkert, N. D. & Lee, J. Using domain adaptation and inductive transfer learning to improve patient outcome prediction in the intensive care unit: Retrospective observational study. *Journal of Medical Internet Research* **26**, e52730 (2024).

[a130] Halloran, K. M. *et al.* Molecular assessment of rejection and injury in lung transplant biopsies. *The Journal of Heart and Lung Transplantation* **38**, 504–513 (2019).

[a131] Fumagalli, I. A. *et al.* Automated ct analysis of body composition as a frailty biomarker in abdominal surgery. *JAMA surgery* **159**, 766–774 (2024).

[a132] Parkes, M. D. *et al.* An integrated molecular diagnostic report for heart transplant biopsies using an ensemble of diagnostic algorithms. *The Journal of Heart and Lung Transplantation* **38**, 636–646 (2019).

[a133] Zhu, J. *et al.* Development and validation of a machine learning algorithm to predict the risk of blood transfusion after total hip replacement in patients with femoral neck fractures: A multicenter retrospective cohort study. *Orthopaedic Surgery* **16**, 2066–2080 (2024).

[a134] Hamilton, E. F. *et al.* Estimating risk of severe neonatal morbidity in preterm births under 32 weeks of gestation. *The Journal of Maternal-Fetal & Neonatal Medicine* **33**, 73–80 (2020).

[a135] Hawken, S. *et al.* External validation of machine learning models including newborn metabolomic markers for postnatal gestational age estimation in east and south-east asian infants. *Gates Open Research* **4**, 164 (2021).

[a136] Yao, M. W. *et al.* Improving ivf utilization with patient-centric artificial intelligence-machine learning (ai/ml): a retrospective multi-center experience. *Journal of Clinical Medicine* **13**, 3560 (2024).

[a137] D'Souza, R. *et al.* Prediction of successful labor induction in persons with a low bishop score using machine learning: Secondary analysis of two randomized controlled trials. *Birth* **50**, 234–243 (2023).

[a138] Ory, J. *et al.* Artificial intelligence based machine learning models predict sperm parameter upgrading after varicocele repair: a multi-institutional analysis. *The world journal of men's health* **40**, 618 (2022).

[a139] Smith, S. C. *et al.* A multicentre development and validation study of a novel lower gastrointestinal bleeding score—the birmingham score. *International journal of colorectal disease* **35**, 285–293 (2020).

[a140] Dunn, W. *et al.* An artificial intelligence-generated model predicts 90-day survival in alcohol-associated hepatitis: A global cohort study. *Hepatology* **80**, 1196–1211 (2024).

[a141] Qi, Y. *et al.* Simulating federated learning for steatosis detection using ultrasound images. *Scientific Reports* **14**, 13253 (2024).

[a142] Guo, L. L. *et al.* A multi-center study on the adaptability of a shared foundation model for electronic health records. *NPJ digital medicine* **7**, 171 (2024).

[a143] Eng, S. W. *et al.* Patterns of joint involvement in juvenile idiopathic arthritis and prediction of disease course: a prospective study with multilayer non-negative matrix factorization. *PLoS medicine* **16**, e1002750 (2019).

[a144] Abdalla, M., Abdalla, M., Hirst, G. & Rudzicz, F. Exploring the privacy-preserving properties of word embeddings: algorithmic validation study. *Journal of medical Internet research* **22**, e18055 (2020).

[a145] Hall, J. N., Galaev, R., Gavrilov, M. & Mondoux, S. Development of a machine learning-based acuity score prediction model for virtual care settings. *BMC Medical Informatics and Decision Making* **23**, 200 (2023).

[a146] Perez-Lebel, A., Varoquaux, G., Le Morvan, M., Josse, J. & Poline, J.-B. Benchmarking missing-values approaches for predictive models on health databases. *GigaScience* **11**, giac013 (2022).

[a147] Cian, L., COVID, I. *et al.* Predicting attitudinal and behavioral responses to covid-19 pandemic using machine learning. *PNAS Nexus, 0 [10.1093/pnasnexus/pgac093]* (2022).

[a148] Wei, J. *et al.* Detecting upper extremity native joint dislocations using deep learning: A multicenter study. *Clinical Imaging* **92**, 38–43 (2022).

[a149] van Spanning, S. H. *et al.* Development and training of a machine learning algorithm to identify patients at risk for recurrence following an arthroscopic bankart repair (clearer): protocol for a retrospective, multicentre, cohort study. *BMJ open* **12**, e055346 (2022).

[a150] Kim, E. H. *et al.* Association of demographic and socioeconomic indicators with the use of wearable devices among children. *JAMA network open* **6**, e235681–e235681 (2023).

[a151] Doshi, S. *et al.* Temporal clustering of critical illness events on medical wards. *JAMA Internal Medicine* **183**, 924–932 (2023).

[a152] Wichmann, B. & Moreira Wichmann, R. Using machine learning to estimate health spillover effects. *The European Journal of Health Economics* **25**, 717–730 (2024).

[a153] Amiri, M. *et al.* Developing and evaluating data infrastructure and implementation tools to support cardiometabolic disease indicator data collection. *Topics in Spinal Cord Injury Rehabilitation* **29**, 124–141 (2023).

[a154] Khera, R. *et al.* Multinational patterns of second line antihyperglycaemic drug initiation across cardiovascular risk groups: federated pharmacoepidemiological evaluation in legend-t2dm. *BMJ medicine* **2**, e000651 (2023).

[a155] Ohayon, M. M., Duhoux, S., Grieco, J. & Côté, M.-L. Prevalence and incidence of narcolepsy symptoms in the us general population. *Sleep Medicine: X* **6**, 100095 (2023).

[a156] Guo, L. L. *et al.* Characterizing the limitations of using diagnosis codes in the context of machine learning for healthcare. *BMC Medical Informatics and Decision Making* **24**, 51 (2024).

[a157] Khera, R. *et al.* Comparative effectiveness of second-line antihyperglycemic agents for cardiovascular outcomes: a multinational, federated analysis of legend-t2dm. *Journal of the American College of Cardiology* **84**, 904–917 (2024).

[a158] Soong, W. *et al.* Disease burden and predictors associated with non-response to antihistamine-based therapy in chronic spontaneous urticaria. *World Allergy Organization Journal* **16**, 100843 (2023).

[a159] Cunningham, K. Y. *et al.* Plasma proteome profiling in giant cell arteritis. *Annals of the rheumatic diseases* **83**, 1762–1772 (2024).

[a160] Nielsen, C., Souza, R., Wilms, M. & Forkert, N. D. Foundation model-driven distributed learning for enhanced retinal age prediction. *Journal of the American Medical Informatics Association* **31**, 2550–2559 (2024).