

Intelligent Clinical Data Platforms for Longitudinal EHR Analytics and Evidence-Based Decision Support

¹Sasi Kumar Kolla, ²Vamsi Krishna Peddi

¹Independent Researcher

sasikkolla@gmail.com

²Independent Researcher

ARTICLE INFO

Received: 02 Aug 2022

Revised 17 Sept 2022

Accepted: 28 Sept 2022

ABSTRACT

Intelligent clinical data platforms enable longitudinal analytics of electronic health records (EHRs), creating chronologically organized datasets for clinical decision-making, patient risk modeling, and biomarkers of disease progression. Well-designed platforms reduce implementation costs and empower clinicians and researchers to address challenges including understanding rare diseases, finding patients for clinical trials, and assessing the impact of interventions by deriving cohorts from EHR data in natural language. Foundational elements required to support these use cases are articulated, including data governance, interoperability, cohort construction, trajectory analysis, and data lineage and quality. EHR datasets typically lack accurate timing information and temporal arrangements, increasing the difficulty of applying risk models, trajectory analysis, and other advanced analytics. High-quality, well-structured electronic Health (eHealth) data promote the construction of clinical decision-support models based on formal evidence derived from real-world data. In the eHealth domain, much effort has been devoted to the development of advanced clinical decision-support models for answering clinically relevant questions, but the preparation and structuring of the required data for advanced analytics have received comparatively little attention. Data-lineage inheritance and quality can hamper the reliability of models built on eHealth data. By providing advanced lineage information through knowledge-generation workflows, auditability and completeness checks enhance the reliability, credibility, and usefulness of generated models. Temporal data representation determines the intrinsic properties of longitudinal datasets and is an essential pillar for longitudinal cohort definition and trajectory modeling. Analytics to support temporal data representation involve the use of methods and techniques for cohort construction and the analysis of cohorts' evolutions over time. Validated temporal patterns can support individual patients in the present and future via real-time analysis and decision support.

Keywords: Intelligent Clinical Data Platforms, Longitudinal Electronic Health Records, Clinical Data Integration, Evidence-Based Decision Support, Predictive Healthcare Analytics, Clinical Data Interoperability, Machine Learning in Healthcare, Patient-Centric Data Analytics, Real-Time Clinical Intelligence, Personalized Clinical Decision Support.

Introduction

Longitudinal EHR Data are Electronic Health Records (EHRs) collected over time from patients within a clinical or healthcare system to support secondary analyses and prognosis [1], [2]. Data in their original, atomic form are stored in a data repository. A clinical data warehouse consolidates data from distinct clinical sources (e.g. the hospital and clinics) for broad data-analysis and supporting services [3]. Intelligent clinical data platforms allow longitudinal EHR Data to serve prospective-analytics required by evidence-based clinical decision support systems that also provide decision-support inputs to healthcare service use and drug-marketing analyses [4], [5]. Formally, each patient's longitudinal record is represented as an ordered sequence of time-stamped events, as defined in Eq. (1). A new approach is presented.

Such data permit a rich set of analyses but many of such analyses may require answers that can be derived from two distinct longitudinal EHR Data constructions: a cohort and a cohort with trajectory

data. Cohorts define patients who either present a type of event or meet other pre-requisite conditions maybe in response to a treatment or syndrome of treatments for a disease, and support a wide range of analytical techniques [6]. A cohort with supporting trajectory data permits visualizing and examining hidden patterns and behaviours leading to patients being in a certain cohort and reveal information not directly pertaining to patients in the cohorts defined [7], [8]. The five key attributes of temporal data representation considered in this work are summarized in Table I.

TABLE I. Five Key Attributes of Temporal Data Representation (Section I-A)

Attribute	Description	Platform handling
Type of time variation	Static, slowly varying, or rapidly varying data	Separate static and time-varying covariate stores
Time-stamping	Whether events carry an explicit time	Mandatory event timestamps; flag missing times
Time intervals	Point events vs. episodes with start/end	Interval encoding $[t_{\text{start}}, t_{\text{end}}]$
Cadence	Regular vs. event-driven sampling	Record the sampling rate per data stream
Irregular observations	Uneven gaps between measurements	Store the gap Δt explicitly (Eq. 2)

A. Temporal data representations

Methods for representing temporal information in clinical databases typically focus on either row-based or temporal graph data models [9]; the two are compared in Table II. The classic tabular model is ill-equipped for representing more complex or temporal aspects of the data. In contrast, a graph-based data representation can naturally include time as an additional data dimension, enabling exploration of complex sequences of for example, medication prescriptions or lab tests [10]. For enabling varied treatments depending on the type of time-varying information being captured in the databases, five key attributes are considered: the type of time variation, whether the data are time-stamped, the time intervals, the type of cadences, and the handling of irregular observations, whose time gaps are defined in Eq. (2). Explicit representation of time in data models can be very beneficial for a number of reasons: it reduces ambiguity in the interpretation of data; it simplifies path-based queries, coding, and auditing; and naturally enables trajectory-based analyses.

In practical terms, modeling time-variant patient health information via temporal representation also supports construction of patient cohorts and subsequent analysis techniques, such as causality assessment or prediction of future time-to-event outcomes [11]. Subsequently, custom temporal representations of different longitudinal health aspects along with model-agnostic analysis methods—including trajectory construction, comparison, clustering, and time-to-event modeling with temporal information as a covariate or factor—can add further capabilities to the databases for specific clinical scenarios.

TABLE II. Row-Based vs. Temporal Graph Data Models (Section I-A)

Criterion	Row-based (tabular)	Temporal graph
Time representation	Implicit column	Explicit dimension on nodes and edges
Event sequences	Needs self-joins	Native path traversal
Path-based queries	Complex	Simple
Auditability	Moderate	High (lineage as edges)
Trajectory analysis	Indirect	Direct

2. Foundations of Longitudinal EHR Data

A foundational requirement for supporting trajectory analyses and evidence-based clinical decision support is the long-term availability of high-quality longitudinal EHR data for a diverse clinical population [12]. Longitudinal analyses aim to derive insights from time-series data over extended periods, while trajectory construction involves identifying, summarizing, and clustering the sequence of relevant states, events, and diagnostic features for a cohort of patients, particularly in the context of complex chronic diseases [13], [14]. Such analyses and constructions depend on pools of past patient data that are sufficiently broad, deep, and unbiased to foster verifiable inferences, predictions, recommendations, and causal implications.

COVID-19 has spotlighted one of the lingering challenges in EHR systems: reliable long-term capture, synthesis, and surfacing of clinically meaningful temporal information [15], [16]. Current solutions fall short of enabling unambiguous analysis, especially when addressing critical challenges such as evaluating treatment effectiveness in the presence of confounding factors, predicting disease trajectories and outcomes while quantifying uncertainty, and inferring the evolutionary paths of infections and their sequelae. A federated intelligence-driven clinical data platform that ensures careful data lineage governance and supports high-quality evidence-based EHR analytics—assessed through a focus on trajectory analysis and cohort construction—holds the potential to solve these long-standing concerns [17]. The data quality dimensions that such a platform must monitor are summarized in Table III, with completeness and timeliness quantified in Eqs. (3) and (4).

TABLE III. Data Quality Dimensions and Example Checks (Section II-A)

Dimension	Definition	Example check
Provenance	History of origin and changes	Lineage record present for each field

Traceability	Link from output back to source	Audit-trail coverage
Completeness	Share of expected values present	q_{comp} (Eq. 3)
Accuracy	Agreement with valid ranges	Lab value within normal range
Consistency	Agreement across sources	Cross-source code reconciliation
Timeliness	Delay between event and record	q_{time} (Eq. 4)

A. Data provenance and quality

The quality of the source data determines the performance of algorithms, machine-learning models, and rule-based systems [18]. A temporary and dynamic set of data quality characteristics serves as quality 'gates' for inbound data flows, formalized as the composite quality gate in Eq. (5). Data from sources used to support decisions, bids, and proposals are typically checked for provenance, traceability, completeness, accuracy, and timeliness. Provenance is the history of data origin and changes over time, including its source, input to processing activities, and lineage. Data quality assessments are performed using test cases and normal ranges for the domain and data type and by analysing audit trails associated with processed data. These are not constrained to automated checks for exactness and validity; far from it. Signature blueprints that record normal behaviour in data augmentation or assurance procedures mapping and tracking entity pair true up have long been built. When all source data are available, business information can be treated like attendance in school, and duplicated the same way, yet with single source data.

Data augmentation adds unobserved descriptions that do not exist in the source data. Patterns or models from conflicting or unlabelled data can be complemented and linked to a known. When an activity or symptom associated with an entity-data pair has been seen repeatedly, it is time to use marker data to complete the part. These performance checks remain a manual activity and are not built into the running system. All work duration numbers spend in the gallery of unused or out-of-use categories remain open to both qualitative and quantitative checking.

Manual quality assurance marked developed and maintained audit trail checks, and tailored business rules are used to provide bias indicator flags. Labelling functions generating the indicator flags and automatically fitting them are under constant improvement with pattern labelling being done visually on the spot where patterns occur. Verification gates for high and low activity yet not yet seen patterns for desire paths on efficiency are now being used to manually brighten marker and activity labels on poster completion.

The sources used to create comment labellers, and detection models for indicator and marker information are tested with various forms of description fusion and threshold tuning, whilst the training patterns are refreshed with added labels, checking progress along the way. A record of the sources of all features of the sources, the quality checks applied to each of the main augmentations in addition to the training and labelling sources used in detection model detection enhances explainability. These quality and lineage controls are embedded in each layer of the platform architecture summarized in Table IV.

TABLE IV. Layers of the Intelligent Platform Architecture (Section III-A)

Layer	Main function	Governance control
Ingestion	API-based intake from EHR, labs, PROs	Consent verification
Storage	Lakehouse and temporal graph	Encryption, access roles
Preprocessing	Harmonization, quality gates	Lineage capture
Analytic	Cohorts and trajectories	Purpose-based access
Processing	Risk and prediction models	Model audit logging
Visualization	Clinician decision support	De-identified views

3. Architectural Principles of Intelligent Clinical Data Platforms

The architectural principles of intelligent clinical data platforms follow the data and services needed for longitudinal EHR analytics. These include data governance to ensure appropriate provenance, usage, and quality of the data; a comprehensive range of analysis services to fully support longitudinal analysis; and intuitive tools for smooth interaction with the platform [19]–[21]. Core analysis services include temporal cohort definition and cumulative-incidence estimation under competing risks, formalized in Eqs. (6) and (7).

The design of intelligent clinical data platforms must consider data provenance, quality, and usage because EHR data are often incomplete and inconsistent. Data Governance within the platform involves specifying the types of operations that can be performed with the data. These capabilities frequently depend on the context of use. For example, using EHR data to estimate the effect of a drug may require demonstrating non-independence-from-ignorance and providing supplemental evidence through instrumental variables or propensity scores [11], computed together with the inverse-probability weights defined in Eq. (8). In contrast, correlation analysis is less restricted. Proper Data Governance can assist with these considerations by regulating data access and use, such as permitting only selected research teams to analyze private patient data, or allowing clinical data scientists to manage the raw data and enable broader use of cleaned data.

The intelligent clinical data platform should therefore monitor provenance, possible bias, and quality, while providing functionalities for cohort construction and trajectory analysis. Cohort Construction is generally offered within a user-friendly interface. Trajectory analysis is included to facilitate the analysis of patient longitudinal data and embrace their temporal dimension. Additionally, allowing patients to record their experiences through Web-based questionnaires can enrich EHR data and extend the range of available analytics [22]. The temporal elements of cohort construction are illustrated in Fig. 1, and Fig. 2 illustrates how trajectory clusters derived from such cohorts can differ in event-free survival.

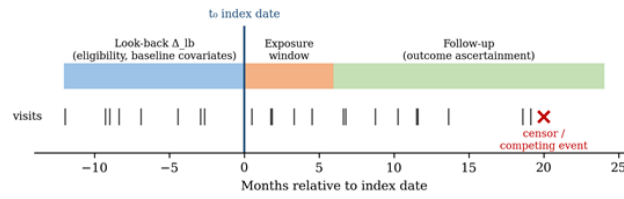


Fig. 1. Temporal cohort construction: look-back window, index date, exposure window, follow-up, and censoring (Section IV-A).

A. Data governance and interoperability

Policies, standards, access controls, consent management, de-identification, and API-based interoperation [9], [17] collectively shape the integrity of CDPs supporting clinical analytics and decision support. These platforms address patient privacy in three ways. Clear policies document patient and institutional rights, particularly concerning sensitive or identifiable information, alongside disclosure of data collection and sharing objectives. Technical methods—data minimization, extraction dependencies, reconciling multi-authority consent—tailor the risk–benefit balance as required. Processes governing which platforms consume which datasets for what reasons acknowledge the tension between open research and specific clinical questions. Enabling patient-centric data sharing across platforms creates richer decision-support solutions while satisfying privacy concerns [23].

An intelligent platform architecture features modular data ingestion, storage, preprocessing, analytic, processing, and visualization layers (Table IV). Each layer can be independently developed, secured, and scaled according to usage. Modular design also increases data provenance, investigation, and auditing capabilities. Control over data access, even among users with analytical permissions, enhances auditing. Carefully monitored logging limits risk, while separation simplifies fault recovery and substitute provisioning. Resilience against errors and adversarial sabotage combines integrity checks, restricted privileges, and separate components.

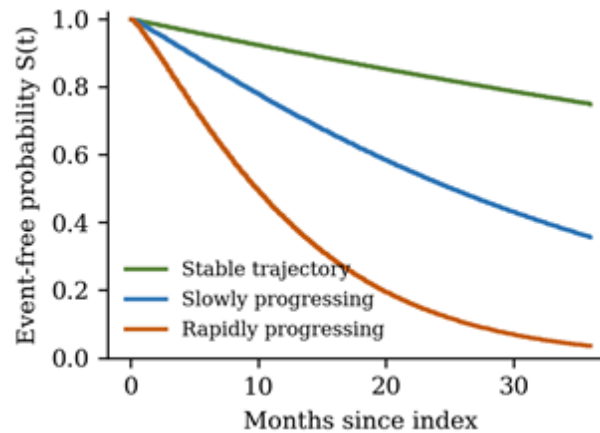


Fig. 2. Illustrative event-free survival curves for three trajectory clusters (synthetic values, for explanation only) (Section IV-A).

4. Analytics for Longitudinal EHR Analytics

Longitudinal analysis of temporal data of patients documented over time requires an intelligent clinical data platform that supports cohort construction and trajectory analysis. Longitudinal studies are among the most powerful nonexperimental study designs, but they typically rely on time-consuming and resource-intensive specialized data collection, making them impractical for many clinical questions. Instead, longitudinal analyses are increasingly performed using routine electronic health record (EHR) data [24], [25]. The absence of bespoke data collections in these studies necessitates evidence-based methods for cohort construction, trajectory modeling, and longitudinal outcome prediction and stratification, as well as approaches for validating and calibrating the resulting models.

Cohort construction and trajectory modeling address two foundations of longitudinal EHR analytics and evidence-based decision support. Methods for cohort construction from current EHR systems are first described, focusing on inclusion criteria tailored for longitudinal studies (e.g., defined exposure windows, months with at least one documented visit), supported by dynamic phenotyping of medications and diseases [13], [14] and time-varying exposure assessments modeled with the time-varying Cox model in Eq. (9). The discussion then turns to the modeling of patient trajectories over time, enrichment of the temporal context with longitudinal outcomes, and development of dynamic risk stratification models that assess the changing short- and long-term risk of predetermined endpoints [26], formalized as the landmark risk in Eq. (10) and illustrated in Fig. 3. This body of work highlights the importance of evidence-based methods in ensuring analytics validity while providing a ready resource for investigators who wish to perform longitudinal studies without incurring the added burden of bespoke data collection.

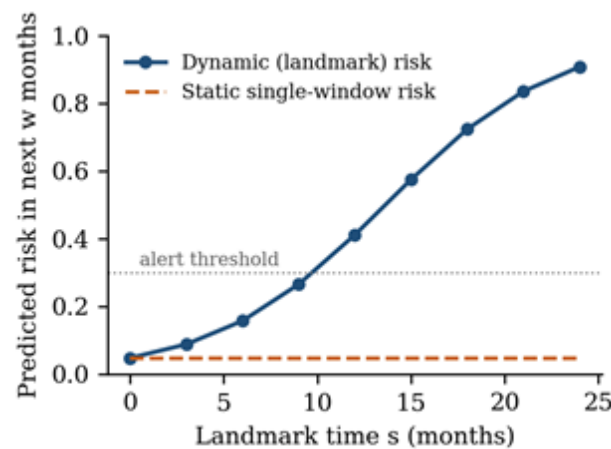


Fig. 3. Illustrative comparison of static single-window risk and dynamic landmark risk over time (synthetic values) (Section IV).

A. Cohort construction and trajectory analysis

Cohorts for trajectory analysis and related longitudinal EHR applications must consider inclusion criteria, target features, temporal characteristics, and suitable analytic methods. Cohorts assure decision-support models are trained and validated on relevant populations, while trajectory analysis identifies discrete patient journeys that may differ with respect to time-to-event outcomes, predictor importance, or hypothetical interventions. Derived trajectories can support an array of applications, covering prognosis, diagnostic reasoning, prediction, and causal inference [27]–[29].

Cohort inclusion criteria typically address demographic and clinical confounders and match the target for an application, but populating the cohort and defining labels for the target feature are not sufficient to support longitudinal pattern analysis. Extracting features for each patient requires temporally-aware approaches, and stratifying patients by event cadence or pattern must consider event types occurring in any order. Cohort analyses typically define a single time window and deploy standard statistical tests, but patient history is seldom stationary, and covariates may be relevant only within discrete time intervals. Analyses that model longitudinal time-varying covariates directly may reduce bias and enhance predictive performance [12], [26]. When an event is known only to lie within an interval, the interval-censored likelihood in Eq. (11) applies, and the calibration of the resulting risk models is assessed with the Brier score in Eq. (12).

Mathematical Formulas:

Longitudinal patient record (Section I) — each patient i is an ordered sequence of events e with time t and value v :

$$P_i = \{(e_{ij}, t_{ij}, v_{ij})\}_{j=1}^{n_i}, \quad t_{i1} < t_{i2} < \dots < t_{in_i} \quad (1)$$

Irregular observation gap (Section I-A):

$$\Delta t_{ij} = t_{ij} - t_{i,j-1} \quad (2)$$

Completeness score (Section II-A):

$$q_{\text{comp}} = \frac{N_{\text{observed}}}{N_{\text{expected}}} \quad (3)$$

Timeliness score — record delay d against a maximum acceptable delay D :

$$q_{\text{time}} = 1 - \frac{1}{N} \sum_{r=1}^N \min \left(\frac{d_r}{D}, 1 \right) \quad (4)$$

Composite quality gate — data are admitted when the weighted score reaches threshold τ :

$$Q = \sum_{k=1}^K w_k q_k, \quad \sum_k w_k = 1, \quad \text{admit if } Q \geq \tau \quad (5)$$

Cohort definition (Section IV-A) — look-back window Δ_{lb} , inclusion set E_{in} , exclusion set E_{ex} :

$$C = \{i: \exists j, e_{ij} \in E_{in}, t_{ij} \in [t_0 - \Delta_{lb}, t_0] \wedge \nexists e_{ij} \in E_{ex}\} \quad (6)$$

Cumulative incidence under competing risks (Section V) — for event type k :

$$F_k(t) = \int_0^t S(u) h_k(u) du \quad (7)$$

Propensity score and inverse-probability weight (Section III) — treatment A , covariates X :

$$e(X) = P(A = 1 | X), \quad w_i = \frac{A_i}{e(X_i)} + \frac{1 - A_i}{1 - e(X_i)} \quad (8)$$

Cox model with time-varying covariates (Section IV-A):

$$h_i(t) = h_0(t) \exp(\beta^T X_i(t)) \quad (9)$$

Dynamic (landmark) risk (Section IV) — risk in the next w months, given survival to landmark s :

$$R_i(s, w) = P(T_i \leq s + w | T_i > s, X_i(s)) \quad (10)$$

Interval-censored likelihood (Section V) — event known to lie in $(L_i, R_i]$:

$$\mathcal{L} = \prod_{i=1}^n [S(L_i) - S(R_i)] \quad (11)$$

Calibration (Brier score) (Section IV):

$$BS = \frac{1}{n} \sum_{i=1}^n (\hat{p}_i - y_i)^2 \quad (12)$$

5. Conclusion

Intelligent clinical data platforms fulfill a clinical data lakehouse's requirement for EHR analytic capabilities while enabling evidence-based decision support by focusing on data provenance, quality, governance, and architecture. Data quality dimensions—completeness, accuracy, consistency, and timeliness—impact downstream analytics. Cohorts and decision-support methods for longitudinal EHR analysis are detailed, including key construction steps, trajectory methods, and statistical models for interval-censored response variables. Building on these foundations, intelligent clinical data platforms have the potential to advance decision-support solutions for clinical care, research, and administration [30].

A demand for evidence-based clinical decision support and medical research is creating interest in the construction of intelligent clinical data platforms: data lakehouses designed to support analytic queries as well as provide high-quality data for predictive models, often supported by an end-to-end Markdown document that documents sources, cleaning steps, and decision-support research datasets. Temporal cohort construction with an emphasis on identifying inclusion/exclusion criteria for cohorts, as well as censoring for competing risks and events not of interest, is one area of clinical data lakehouse analytics. Methods for the identification of longitudinal temporal trends and activity churn in populations are essential capabilities.

References

- [1] Bastarache, L., Brown, J. S., Cimino, J. J., Dorr, D. A., Embi, P. J., Payne, P. R. O., Wilcox, A. B., & Weiner, M. G. (2022). Developing real-world evidence from real-world data: Transforming raw data into analytical datasets. *Learning Health Systems*, 6(1), e10293.
- [2] Wells, B. J., Downs, S. M., & Ostasiewski, B. (2022). Using electronic health records for the learning health system: Creation of a diabetes research registry. *JMIR Medical Informatics*, 10(9), e39746.

- [3] Lim, H. C., Austin, J. A., van der Vegt, A. H., Dumble, E., Martin, B., & Sullivan, C. (2022). Toward a learning health care system: A systematic review and evidence-based conceptual framework for implementation of clinical analytics in a digital hospital. *Applied Clinical Informatics*, 13(2), 339–354.
- [4] Hak, F., Guimarães, T., & Santos, M. (2022). Towards effective clinical decision support systems: A systematic review. *PLOS ONE*, 17(8), e0272846.
- [5] Balasubramanya, B., Coffey, P., Abeyaratne, A., & Cass, A. (2022). Barriers and enablers to implementing and using clinical decision support systems for chronic diseases: A qualitative systematic review and meta-aggregation. *Implementation Science Communications*, 3, 81.
- [6] Brnabic, A., & Hess, L. M. (2021). Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making. *BMC Medical Informatics and Decision Making*, 21, 54.
- [7] Kaplan, A. D., Greene, J. D., Liu, V. X., & Ray, P. (2022). Unsupervised probabilistic models for sequential electronic health records. *Journal of Biomedical Informatics*, 134, 104163.
- [8] Kaplan, A. D., Tipnis, U., Beckham, J. C., Kimbrel, N. A., Oslin, D. W., McMahon, B. H., & MVP Suicide Exemplar Workgroup. (2022). Continuous-time probabilistic models for longitudinal electronic health records. *Journal of Biomedical Informatics*, 130, 104084.
- [9] Xiao, G., Pfaff, E., Prud'hommeaux, E., Booth, D., Sharma, D. K., Huo, N., Yu, Y., Zong, N., Ruddy, K. J., Chute, C. G., & Jiang, G. (2022). FHIR-Ontop-OMOP: Building clinical knowledge graphs in FHIR RDF with the OMOP Common Data Model. *Journal of Biomedical Informatics*, 134, 104201.
- [10] Xie, F., Yuan, H., Ning, Y., Ong, M. E. H., Feng, M., Hsu, W., Chakraborty, B., & Liu, N. (2022). Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies. *Journal of Biomedical Informatics*, 126, 103980.
- [11] Ding, P., Pan, Y., Wang, Q., & Xu, R. (2022). Prediction and evaluation of combination pharmacotherapy using natural language processing, machine learning and patient electronic health records. *Journal of Biomedical Informatics*, 133, 104164.
- [12] Jin, Y., Weberpals, J. G., Wang, S. V., Desai, R. J., Merola, D., & Lin, K. J. (2022). The impact of longitudinal data-completeness of electronic health record data on the prediction performance of clinical risk scores. *Clinical Pharmacology & Therapeutics*, 113(6), 1359–1367.
- [13] Ahuja, Y., Zou, Y., Verma, A., Buckeridge, D., & Li, Y. (2022). MixEHR-Guided: A guided multi-modal topic modeling approach for large-scale automatic phenotyping using the electronic health record. *Journal of Biomedical Informatics*, 134, 104190.
- [14] Nogues, I.-E., Wen, J., Lin, Y., Liu, M., Tedeschi, S. K., Geva, A., Cai, T., & Hong, C. (2022). Weakly semi-supervised phenotyping using electronic health records. *Journal of Biomedical Informatics*, 134, 104175.
- [15] Li, I., Pan, J., Goldwasser, J., Verma, N., Wong, W. P., Nuzumlal, M. Y., Rosand, B., Li, Y., Zhang, M., Chang, D., Taylor, R. A., Krumholz, H. M., & Radev, D. (2022). Neural natural language processing for unstructured data in electronic health records: A review. *Computer Science Review*, 46, 100511.
- [16] Liu, J., Capurro, D., Nguyen, A., & Verspoor, K. (2022). “Note bloat” impacts deep learning-based NLP models for clinical prediction tasks. *Journal of Biomedical Informatics*, 133, 104149.
- [17] Yoo, J., Lee, J., Min, J. Y., Choi, S. W., Kwon, J.-M., Cho, I., Lim, C., Choi, M. Y., & Cha, W. C. (2022). Development of an interoperable and easily transferable clinical decision support system deployment platform: System design and development study. *Journal of Medical Internet Research*, 24(7), e37928.
- [18] van Os, H. J. A., Kanning, J. P., Wermer, M. J. H., Chavannes, N. H., Numans, M. E., Ruigrok, Y. M., van Zwet, E. W., Putter, H., Steyerberg, E. W., & Groenwold, R. H. H. (2022). Developing clinical prediction models using primary care electronic health record data: The impact of data preparation choices on model performance. *Frontiers in Epidemiology*, 2, 871630.

- [19] Foley, T., & Vale, L. (2022). A framework for understanding, designing, developing and evaluating learning health systems. *Learning Health Systems*, 7(1), e10315.
- [20] Friedman, C. P. (2022). What is unique about learning health systems? *Learning Health Systems*, 6(3), e10328.
- [21] Azar, K. M. J., Pletcher, M. J., Greene, S. M., & Pressman, A. R. (2022). Learning health system, positive deviance analysis, and electronic health records: Synergy for a learning health system. *Learning Health Systems*, 7(3), e10348.
- [22] Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F., & Luo, Y. (2022). Multimodal machine learning in precision health: A scoping review. *npj Digital Medicine*, 5, 171.
- [23] Stipelman, C. H., Kukhareva, P. V., Trepman, E., Nguyen, Q.-T., Valdez, L., Kenost, C., Hightower, M., & Kawamoto, K. (2022). Electronic health record-integrated clinical decision support for clinicians serving populations facing health care disparities: Literature review. *Yearbook of Medical Informatics*, 31(1), 184–198.
- [24] Jabbour, S., Fouhey, D., Kazerooni, E., Wiens, J., & Sjoding, M. W. (2022). Combining chest X-rays and electronic health record data using machine learning to diagnose acute respiratory failure. *Journal of the American Medical Informatics Association*, 29(6), 1060–1068.
- [25] Abraham, A., Le, B., Kosti, I., Straub, P., Velez-Edwards, D. R., Davis, L. K., Newton, J. M., Muglia, L. J., Rokas, A., Bejan, C. A., Sirota, M., & Capra, J. A. (2022). Dense phenotyping from electronic health records enables machine learning-based prediction of preterm birth. *BMC Medicine*, 20, 333.
- [26] Mamouei, M., Budhdeo, S., Seth, S., Hall, P., & Elliott, P. (2022). Validation of risk prediction models applied to longitudinal electronic health record data for the prediction of major cardiovascular events in the presence of data shifts. *European Heart Journal – Digital Health*, 3(4), 535–547.
- [27] Kim, J. H., Goldberg, T. E., & Weng, C. (2022). A machine learning approach to identifying delirium from electronic health records. *JAMIA Open*, 5(2), ooac042.
- [28] Liu, S., Schlesinger, J. J., McCoy, A. B., Reese, T. J., Steitz, B., Russo, E., Koh, B., & Wright, A. (2022). New onset delirium prediction using machine learning and long short-term memory in electronic health record. *Journal of the American Medical Informatics Association*, 30(1), 120–131.
- [29] Hong, N., Liu, C., Gao, J., Han, L., Chang, F., Gong, M., & Su, L. (2022). State of the art of machine learning-enabled clinical decision support in intensive care units: Literature review. *JMIR Medical Informatics*, 10(3), e28781.
- [30] Kouri, A., Yamada, J., Lam Shin Cheung, J., & Van de Velde, S. (2022). Do providers use computerized clinical decision support systems? A systematic review and meta-regression of clinical decision support uptake. *Implementation Science*, 17, 21.