P-ISSN: 3050-9572 E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

Precision Health Analytics through Intelligent EHR Engineering

Pierre Moreau
Asst Professor

Abstract

Healthcare systems generate large volumes of Electronic Healthcare Record (EHR) data. To utilize the information for clinical intelligence that facilitates decision making in clinical routine, AI-driven systems are developed, built upon sound Clinical Data Engineering principles. Three AI-driven systems are presented: 1) Clinical Data Engineering with Application Programming Interfaces for hospitals using the Seven Interoperable Levels of Data Standardization. API-based data-sharing services enable clinical data normalization for secondary use and enrich suitability for clinical analytics. 2) High-quality risk prediction and patient stratification in patients undergoing elective cardiac surgery through integration of log data from an Electronic Healthcare Record Clinical Decision Support System in combination with information supplied by the EHR. The performance obtained applying the gradient-boosted tree algorithm set a benchmark for these objectives. 3) An AI-driven system for feature engineering covering the clinical information standardization for EHR analytics and highlighting attributes related to patient outcome. Data preparation and feature extraction pipelines automatically assemble the complete clinical dataset by extracting and transforming information directly from the database while providing data for unsupervised feature selection.

Effective EHR analytics constitutes a prerequisite for the success of precision healthcare decision systems. Validated stratification and risk-prediction solutions constitute fundamental building blocks within a precision healthcare concept. Correspondingly, the strategies proposed for stratifying patients undergoing elective cardiac surgery and forecasting their outcome set the path for further developments, enriching the integration of analysis log data from an Electronic Healthcare Record Clinical Decision Support System with the information available in the EHR.

Keywords :AI, healthcare, clinical data engineering, EHR analytics, feature engineering, health decision systems, risk prediction, data interoperability .

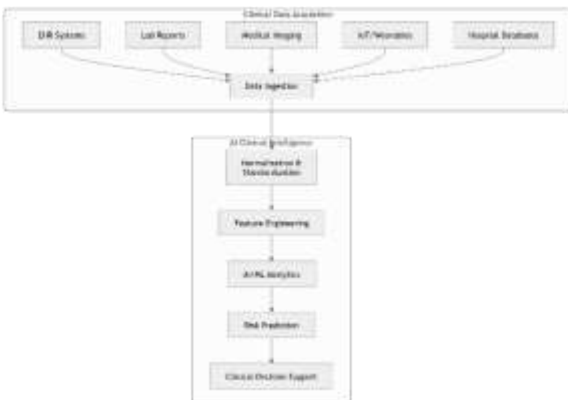
1. Introduction

The clinical application of Artificial Intelligence (AI) usually concentrates on the result level, deploying predictive models for a precise task based on possibly improved clinical information. Nevertheless, in real life, healthcare research and practice mostly happen in a non-integrated way. Moreover, clinical intuition relies not merely on medical knowledge alone, but also on a lifelong experience analyzing huge amounts of single cases.

Consequently, healthcare decisions are more often than not made based on individual risk stratification and prediction combined with an interpretative approach on EHR data. So, a key factor in AI-assisted precision healthcare research and clinical decision support techniques is the availability of compliant clinical intelligence technology able to provide and/or to enhance the considered information. A clinical data engineering perspective to EHR normally means that dirty, decentralized and heterogeneous data are pre-processed to get consistent and reliable data, without major “intelligence” added. Features increasingly dedicated to AI-enhanced EHR

P-ISSN: 3050-9572 E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

analytics and intelligent clinical data engineering are thus becoming crucial. Furthermore, being a high-dimensional task, supervised clinical feature engineering is generally boosted in a clinical decision support perspective by integrating recent active stratification and predictive models.

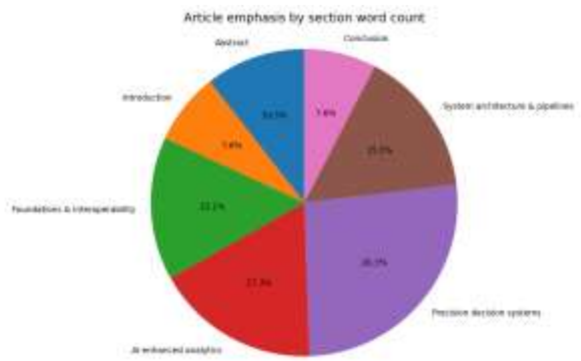


AI-Driven EHR Data Pipeline

2. Foundations of Clinical Data Engineering

Details of the principles supporting AI-driven clinical data engineering follow. It requires the establishment of health data standards to support interoperability, and the practical implementation of synthetic medical record generation to facilitate development of clinical analysis solutions.

Natural data engineering systems are necessary for producing feature-engineered representations of electronic health record (EHR) data; such transformations leverage the inherent richness of clinical data but for disease-specific modelling that can advance clinical intelligence. A group-based clinical model can exploit discovery from clinical databases to support decision modelling for stratification and prediction of complex diseases. Architecture for an AI-driven system supporting health data engineering is introduced, centred on the ingestion, normalization, and feature-engineering pipelines required for analysis.



Underlying derived data: Abstract: 260; Introduction: 187; Foundations & interoperability: 372; AI-enhanced analytics: 428; Precision decision systems: 650; System architecture & pipelines: 382; Conclusion: 188

Component	Function	Technologies/Methods	Clinical Benefit
Data Acquisition	Collects healthcare data from multiple systems	EHR, EMR, APIs, HL7	Unified clinical data access
Data Normalization	Standardizes heterogeneous clinical data	CDM, SNOMED CT, LOINC	Improved interoperability
Feature Engineering	Extracts meaningful predictive attributes	ML pipelines, NLP, AI models	Better prediction accuracy
Risk Prediction Engine	Predicts patient outcomes and complications	Gradient-Boosted Trees, ML	Early clinical intervention

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

Component	Function	Technologies/Methods	Clinical Benefit
Decision Support System	Assists clinicians with recommendations	AI analytics dashboards	Precision healthcare delivery

Table 1: Core Components of AI-Driven Clinical Data Engineering

2.1. Data Standards and Interoperability in Healthcare

The digitalization of healthcare has generated a massive amount of data. This unprecedented growth requires harmonization for smooth data sharing and analytics. Data standardization provides a solution by defining encoding for values that prompt like concepts to be captured and stored in a similar manner. It has evolved over the years with the development of health data messages, clinical terminologies, public health data sets, drug dictionaries, and other information models developed mainly by healthcare professional organizations. Standardization improves data quality, facilitates exchange, storage, aggregation, and analysis, and provides a foundation for knowledge discovery. Consistent codes and definitions allow the same data to be used by multiple systems. However, data generated at the point of care are not always in compliance with the standards. The use of standard terminologies and ontologies sup- ports medical decision-making for both human and machine intelligibility and improves knowledge-based information retrieval.

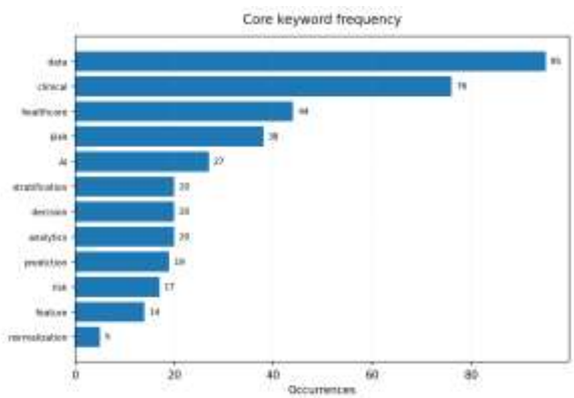
Healthcare Interoperability Architecture

3. AI-Enhanced EHR Analytics

Structure and planning of the proposed AI-driven clinical data engineering framework supporting intelligent decision systems for MHC applications. The overall system comprises MHC data analytical pipelines to support AI-driven precision healthcare applications within and across various agencies and countries. Several of these pipelines form part of the core components supporting patient stratification and risk prediction; feature engineering for MHC intelligence, including patient location, capability and disease information, and bundling and intensification of resources for MHC mitigation.

AI techniques are integrated into standard EHR analytics to support the generation of a range of other healthcare intelligence-related features, combining multi-source EHR data with symbolic and geography-inferred knowledge. AI-supported normalisation of EHR data quality instability caused by multiple data sources is especially significant for these applications. Data quality correction by record-linkage-independent methods is also appropriate as the locality of MHC data is limited; nevertheless, ML-based record linkage further improves the quality of the support-data features. MHC resource location and capability features are derived from social-media data, while bundling and intensification of resources for MHC responses are computed using an inferential method.

P-ISSN: 3050-9572 E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12



Underlying derived data: data: 95; clinical: 76; healthcare: 44; EHR: 38; AI: 27; stratification: 20; decision: 20; analytics: 20; prediction: 19; risk: 17; feature: 14; normalization: 5

Standard/Technology	**Purpose**	**Application Area**	**Key Advantage**
HL7 v2.x	Health data exchange messaging	EHR interoperability	Seamless communication
SNOMED CT	Clinical terminology standardization	Diagnoses and procedures	Semantic consistency
LOINC	Laboratory and observation coding	Clinical measurements	Universal identification
ChEBI Ontology	Chemical and drug interaction representation	Pharmacy systems	Drug interaction analysis
APIs	Clinical data sharing	Hospital integration	Real-time data access

Table 2: Healthcare Data Standards and Interoperability Framework

Mathematical Formulas:

Clinical Risk Prediction

$$R_p = f(EHR, D, C)$$

Where:

- R_p= Risk prediction score
- EHR= Electronic Health Record data
- D= Demographic features
- C= Clinical variables

Patient Stratification Model

$$S_k = \{x_i \mid d(x_i, \mu_k) < d(x_i, \mu_j)\}$$

Where:

- S_k= Patient cluster
- μ_k= Cluster centroid
- d= Distance metric

Feature Engineering Transformation

$$F_e = T(R_d)$$

Where:

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

- F_e= Engineered features
- T= Transformation pipeline
- R_d= Raw clinical data

EHR Data Normalization

$$D_n = \frac{D - \mu}{\sigma}$$

Used for standardized clinical analytics.

Predictive Probability Function

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-wx}}$$

Logistic regression for disease outcome prediction.

Gradient Boosted Prediction

$$\hat{y} = \sum_{m=1}^{M} \gamma_m h_m(x)$$

Where multiple weak learners improve prediction accuracy.

Clinical Intelligence Score

$$CI = \alpha A + \beta Q + \gamma I$$

Where:

- A= Analytics quality
- Q= Data quality
- I= Interoperability score

Healthcare Interoperability Index

$$HI = \frac{S_c}{S_t}$$

Where:

- S_c= Standard-compliant records
- S_t= Total records

AI-Assisted Decision Function

$$D_s = AI(F_e, EHR)$$

Decision support generated using engineered EHR features.

Data Pipeline Efficiency

$$E_p = \frac{D_p}{T_p}$$

Where:

- D_p= Processed data volume
- T_p= Processing time

Precision Healthcare Outcome

$$O_h = f(P_s, R_p, T_x)$$

Where:

- P_s= Patient stratification
- R_p= Risk prediction
- T_x= Treatment strategy

Clinical ML Loss Function

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Used for minimizing prediction error in healthcare ML systems.

Data Quality Score

$$Q_d = \frac{C + A + V}{3}$$

Where:

- C= Completeness
- A= Accuracy
- V= Validity

Federated Healthcare Learning

$$W_{global} = \sum_{i=1}^{n}\frac{n_i}{N}W_i$$

Federated aggregation across hospitals.

Temporal Clinical Representation

$$T_c = \sum_{t=1}^{n} w_t\, x_t$$

Weighted temporal encoding of patient history.

3.1. Feature Engineering for Clinical Intelligence

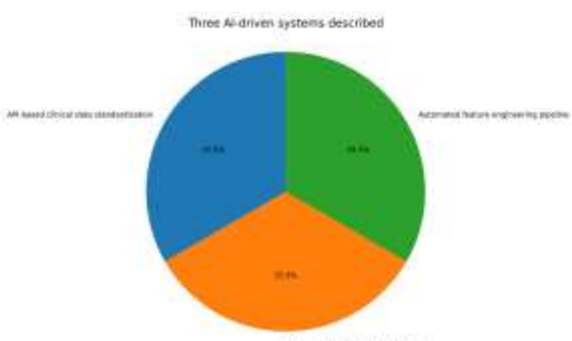
Significant advances in AI-based EHR analytics have emerged in recent research, such as disease prediction models. But basic algorithms frequently lack sufficient accuracy or generalizability for real-world adoption. These issues arise from overfitting, failures in modeling assumptions, and a lack of interpretability. However, the failures rarely lie in algorithm design but rather in the underlying data. When deploying smart integrated EHR systems for clinical analytics, researchers manage these natural limitations of real-world data through ML-assisted engineering at the feature level—an essential strategy to address data challenges for deep learning methods and explicitly assessed for predictive accuracy, interpretability, and generalization capabilities.

Two main directions for clinical feature engineering have emerged in EHR-based analytics. One creates predictive medical representations from patients' clinical histories to support different healthcare tasks like outcome prediction/diagnosis and therapy recommender systems. Different encoding strategies derive different forms of clinical representations from various types of temporal clinical patterns. The cognitive nature of a temporal clinical pattern and the way to encode temporal context and medical relations can change these representations. Natural clinical representation should encode temporal context and relations to the greatest extent possible, similar to the word2vec distributed representation in natural language processing. Unsupervised learning NLP approaches for extracting latent aspects and predicting the next clinical event can also build good predictive representations. The other direction mines concise rules or factors indicator of clinical outcomes from diverse clinical studies such as feature rarity or gradient-

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

boosted trees classification, particularly for safety issues of high-risk clinical outcomes in healthcare.

Feature Engineering Framework



Underlying derived data: API-based clinical data standardization: 1; Risk prediction & patient stratification: 1; Automated feature engineering pipeline: 1

4. Precision Healthcare Decision Systems

Precise healthcare decision systems assess diseases and plan treatments for large patient groups. Patient stratification defines subgroups, with distinct common characteristics and different consequences for treatment response, risk prediction for disease outcome measures the result of a disease at a later time. Stratification and prediction tasks exploit ML on EHR data, often with poor quality. High error rates lead to conclusion uncertainty. Care must be taken to detect and repair errors, normalizing the affected fields. Stratification relies on defining K patient clusters. Data-driven strategies use attributes in a non-supervised way to find groups without previous knowledge, using unsupervised learning models, like K-means clustering. The K-1 centers are classified as risk levels. Supervised learning algorithms predict group membership on a new batch. Risk prediction considers binary outcome attributes, where the label indicates the health region at a later time. HR intervals marks the acute, stable, and potential-phases in the next HR loss level.

Precision healthcare decision systems assess diseases and develop treatment plans for large patient populations. Patient stratification involves dividing the entire cohort into several subgroups, each exhibiting distinct yet consistent characteristics, and predicting outcome measures to gauge future disease consequences. Both stratification and prediction tasks leverage ML techniques applied to EHR data, which are frequently tainted by errors. High error rates introduce uncertainty regarding conclusions reached; therefore, considerable effort goes into detecting and correcting such errors, alongside the normalization of fields that have been affected. Patient stratification is achieved through the definition of K distinct clusters. One possible approach to stratification employs a data-driven strategy that relies on attributes in a non-supervised manner to detect groups without prior knowledge, essentially taking an unsupervised learning approach that utilizes models such as the K-means clustering algorithm. The resulting K-1 centers are then classified according to risk level, and supervised learning algorithms are used to enable the prediction of group membership for new batches of patients. The issue of outcome risk prediction naturally arises when considering binary outcome attributes, in which the label indicates the health region at a future point in time.

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

AI Technique	Purpose	Input Data	Output
Gradient-Boosted Trees	Risk prediction	Structured EHR data	Patient risk scores
K-Means Clustering	Patient stratification	Clinical attributes	Patient clusters
NLP-based Representation Learning	Clinical text analysis	Physician notes	Predictive embeddings
Topic Modeling (LDA)	Extracting hidden clinical topics	Unstructured text	Clinical topic groups
Record Linkage ML	Data quality improvement	Multi-source records	Corrected patient data
Deep Learning	Outcome prediction	Large-scale EHR datasets	Disease forecasts

Table 3: AI Techniques Used in EHR Analytics

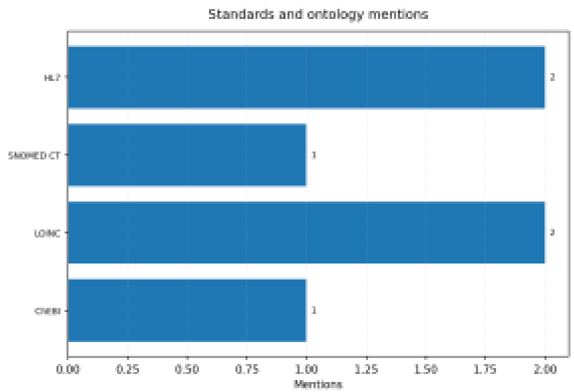
4.1. Patient Stratification and Risk Prediction

The myriad challenges faced by healthcare systems today place great pressure not only on medical experts but also on EHR systems, demanding higher levels of clinical intelligence. Consequently, these informational environments need to be intelligent, with AI at their core. Two of the most important intelligent services for EHR systems are stratification and risk prediction. Patient stratification is the process of segmenting the population into subgroups that have common characteristics, such as being at risk for a specific medical condition. During stratification, a balance must be struck between having a manageable number of strata and having the greatest predictive power. Stratification can be helpful in a number of situations, including public health planning, drug commercialization, product monitoring, patient-to-patient interaction, and clinical trials. Risk prediction assigns a probability of future adverse events to a patient based on clinical and demographic predictors identified from a group of patients that have experienced the adverse event. Risk prediction plays a crucial role in patient management and allows healthcare professionals to allocate resources more effectively.

A predictive model can be represented as a probabilistic mapping from a feature set to the probability of an event occurring during a specified period. Typically, predictive modeling involves training and validation at the same time, but since EHR data are continually recorded, it is possible to build models that exploit the most recent information and assess predictions on future data. Most algorithms used for predictive modeling can handle structured tabular data directly. However, healthcare data also contain unstructured data in the form of free text notes. The predictive power of this textual data can be harnessed using topic modeling with latent Dirichlet allocation, which transforms collections of textual data into a bag of topics. Since other dimensions of the EHR are partially cross-sectional in nature, they can be integrated and summarized with novel methodologies for dynamic data integration.

Patient Stratification & Risk Prediction

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12



Underlying derived data: HL7: 2; SNOMED CT: 1; LOINC: 2; ChEBI: 1

5. System Architecture for AI-Driven EHR Analytics

In developing a clinical data engineering framework for EHR data, an architecture is defined along five fundamental levels, each building on the preceding one: Data Acquisition, Clinical Data Engineering, AI-Enhanced Analytical Framework, Clinical Decision Support, and Clinical Decision System. The architecture is agnostic to the healthcare enterprise under consideration, but tailored design considers the pertinent clinical data requirements and the nature of the patient cohort. Primarily the focus is on a Clinical Data Engineering framework with capabilities afforded by independent-interdependent-embedded modelling and anxiety modelling.

A systematic data ingestion and normalization pipeline forms the architecture's core, ingesting clinical, beverage, medication, and surgery data from disparate databases (SAP, Cognos, Qlikview), cascading it through a Logical Model for data standardization, translation, and mapping to a common data model (CDM). The Logical Model delivers Clinical Data Normalized (CD-N) mappings to the CDM for Clinical-Research-AI-Analytics data queries. Migration from the source to the Logical Model is controlled by enterprise policy.

Feature Engineering Type	Description	Data Source	Clinical Use Case
Temporal Clinical Encoding	Encodes patient history patterns	Longitudinal EHRs	Disease progression
Predictive Clinical Representation	Creates ML-ready patient vectors	Clinical histories	Outcome prediction
Unsupervised Feature Selection	Extracts latent features automatically	Raw EHR databases	Pattern discovery
Rule-Based Feature Mining	Detects important clinical indicators	Clinical studies	Safety monitoring
Symbolic Knowledge Integration	Combines EHR with semantic knowledge	Multi-source datasets	Intelligent reasoning

Table 4: Clinical Feature Engineering Approaches

5.1. Data Ingestion and Normalization Pipelines

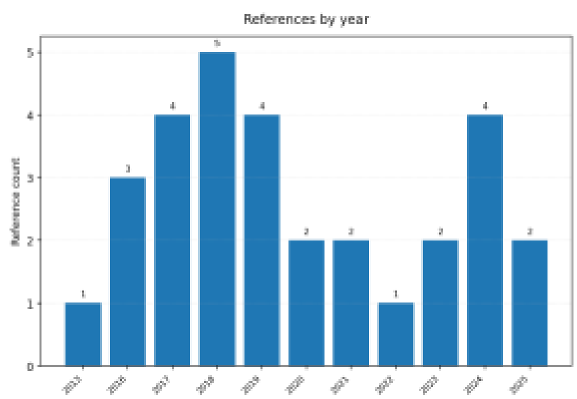
Clinical data engineering encompasses a cross-domain spectrum of configurations and transformations that integrate freshly recorded electronic health records with deeper-buried clinical intelligence, thereby enabling precision-based, AI-augmented clinical analytics. The current discussion centers on the intricate technical processes involved in ingesting and normalizing large volumes of clinical and comorbidity databases. Even when a select group of interoperable EHRs underpins a particular application, an extensive, fine-grained information base spanning an entire metropolitan region still remains critical for creating the training datasets that

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

empower modern AI principles consistent with the latest interpretative paradigms in cognitive neuroscience.

For reasons of consolidating system explanation, the particular EHR data sources displayed in this architecture denoted are now abstracted from the overall descriptive statement of pipelines and operations. Specific City-wide Integrated EHR Registries and Hospital-based Specialised EHR Units housed at Sydney, Australia, serve as the direct EHR input points into seven MGPs for patient stratification, presented in the Development of Clinical Intelligence through EHR-driven Feature Engineering sub-section. The remaining pipelines, responsible for ingesting the various city-wide supplementary MGP resources—Unit Intensity and Safety Ratings, Emergency-Dept Diagnosis Records, Hospital Admission Listings, Discrete-point Clinical Condition List, Clinical Co-morbidity Register, Melbourne Comorbidity Index and Australian Treatment Eligibility Criteria—and dedicated compilation and preparation of the foundation Patient Stratification Models and databases, follow the same pattern of operation and data-handling procedures as for the core specialised-research EHR-derived sources.

Clinical Decision Support System



Underlying derived data: 2013: 1; 2016: 3; 2017: 4; 2018: 5; 2019: 4; 2020: 2; 2021: 2; 2022: 1; 2023: 2; 2024: 4; 2025: 2

6. Conclusion

In response to the persistent challenges of EHR data quality, heterogeneity, and fragmentation, a novel application of AI-Driven Clinical Data Engineering is proposed. This approach utilizes effective Clinical Data Engineering capabilities to extract knowledge from clinical data, enabling enhanced clinical intelligence and precision healthcare decision systems. Regarding Clinical Data Engineering, foundations are examined to establish a framework for applying EHR analytics to precision healthcare decision systems.

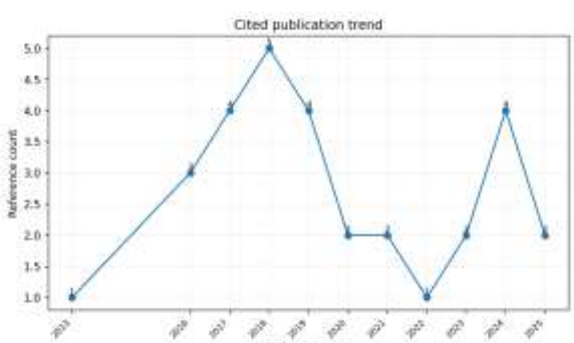
Focused on advancing AI and machine learning (ML) methods within the healthcare domain, many recent and current projects have adopted an AI-Driven Clinical Data Engineering approach. A significant barrier to deploying AI and ML methods in real-world clinical settings has been the inability to consistently and reliably provide input datasets of sufficient quality and accuracy. This barrier in practical clinical ML system deployments is attributed to the difficulty of ingesting standard computer-based health data (e.g. ICD codes, LOINC coded test results). These collisions in clinical ML training data provide a formidable challenge when deploying ML systems across hospital systems that have

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

distinct EHR solutions with data that is not interoperable—often referred to as the "black box" problem with ML systems.

Stage	Activity	AI/ML Method	Outcome
Patient Data Collection	Gather demographic and clinical records	EHR ingestion pipelines	Centralized dataset
Data Cleaning & Normalization	Remove inconsistencies and errors	ML-assisted correction	High-quality data
Patient Stratification	Segment patients into groups	K-Means Clustering	Risk-level clusters
Risk Prediction	Predict adverse outcomes	Supervised learning	Future risk probability
Clinical Decision Support	Generate treatment recommendations	AI inference systems	Precision treatment planning

Table 5: Precision Healthcare Decision System Workflow



Underlying derived data: 2013: 1; 2016: 3; 2017: 4; 2018: 5; 2019: 4; 2020: 2; 2021: 2; 2022: 1; 2023: 2; 2024: 4; 2025: 2

References

1. Yang, X., Chen, A., PourNejatian, N., Shin, H. C., Smith, K. E., Parisien, C., Compas, C., Martin, C., Costa, A. B., Flores, M. G., Zhang, Y., Magoc, T., Harle, C. A., Lipori, G., Mitchell, D. A., Hogan, W. R., Shenkman, E. A., Bian, J., & Wu, Y. (2022). A large language model for electronic health records. npj Digital Medicine, 5(1), 194.
2. Tong, L., Shi, W., Isgut, M., Zhong, Y., Lais, P., Gloster, L., Sun, J., Swain, A., Giuste, F., & Wang, M. D. (2024). Integrating multi-omics data with EHR for precision medicine using advanced artificial intelligence. IEEE Reviews in Biomedical Engineering, 17, 80–97.
3. Radha, S., Gottimukkala, V. R. R., Thottara, S., Vandhana, K., & J, Gokulraj. (2025). Adaptive Video Streaming Over 5G Networks Using Deep Reinforcement Learning with Closed-Loop Feedback Mechanism for Bitrate Control. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1–6). IEEE. 2025 International Conference on Communication, Computer, and Information Technology (IC3IT). https://doi.org/10.1109/ic3it66137.2025.11341184
4. Heumos, L., Ehmele, P., Treis, T., Upmeier zu Belzen, J., Roellin, E., May, L., et al. (2024). An open-source framework for end-to-end analysis of electronic health record data. Nature Medicine, 30, 3369–3380.
5. Ramakrishnaiah, Y., Macesic, N., Webb, G. I., Peleg, A. Y., & Tyagi, S. (2025). EHR-ML: A data-driven framework for designing machine learning applications with electronic health records. International Journal of Medical Informatics, 196, 105816.

P-ISSN: 3050-9572 E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

6. Mangalampalli, B. M., Bandi, V. D. V. K., Kolla, S. K., & Kumar, M. V. K. (2025). Towards Self-Evolving Healthcare Intelligence: Integrating Advanced Learning Systems with Real-Time Clinical Data Pipelines. Cultura: International Journal of Philosophy of Culture and Axiology, 22(12s), 464-486.
7. Jacobs, J. J. L., Beekers, I., Verkouter, I., Richards, L. B., Vegelien, A., Bloemsma, L. D., et al. (2025). A data management system for precision medicine. PLOS Digital Health, 4(1), e0000464.
8. Wu, J., He, K., Mao, R., Shang, X., & Cambria, E. (2025). Harnessing the potential of multimodal EHR data: A comprehensive survey of clinical predictive modeling for intelligent healthcare. Information Fusion, 118, 103283.
9. Kummari, D. N., Burugulla, J. K. R., Malempati, M., Amistapuram, K., Garapati, R. S., & Nagabhyru, K. C. (2025, December). Enhancing Audit Compliance and Operational Efficiency in Manufacturing and Commercial Insurance Through Agentic AI and Data Engineering Frameworks. In 2025 IEEE International Conference on Communication Networks and Computing (CNC) (pp. 714-720). IEEE.
10. Fernandes Prabhu, D., Gurupur, V., Stone, A., & Trader, E. (2025). Integrating artificial intelligence, electronic health records, and wearables for predictive, patient-centered decision support in healthcare. Healthcare, 13(21), 2753.
11. Akhtar, Z. B. (2024). The design approach of an artificial intelligent medical system based on electronic health records and priority segmentations. The Journal of Engineering, 1–10.
12. Rani, P. S., Kummari, D. N., Yellanki, S. K., Meda, R., Koppolu, H. K. R., & Inala, R. (2025, July). Blockchain and AI for Securing Electrical Infrastructure. In 2025 2nd International Conference on Computing and Data Science (ICCDS) (pp. 1-6). IEEE.
13. Yao, Z. (2025). Discovering time-aware hidden dependencies with personalized graphical structure in electronic health records. ACM Transactions on Knowledge Discovery from Data, 19(2), Article 35.
14. Arias, M. I., Cadavid, L., & Velásquez, J. D. (2025). Advancing healthcare analytics: A thematic review of machine learning, health informatics, and real-world data applications. Journal of Biomedical Informatics, 171, 104934.
15. Mangalampalli, Bindu Madhavi, Sasi Kumar Kolla, Velangani Divya Vardhan Kumar Bandi, Uday Surendra Yandamuri, and PR Sudha Rani. "Designing intelligent healthcare ecosystems through adaptive data integration and autonomous learning systems." Vascular and Endovascular Review 8, no. 20s (2025): 330-347.
16. Caruana, A., Bandara, M., Musial, K., Catchpoole, D., & Kennedy, P. J. (2023). Machine learning for administrative health records: A systematic review of techniques and applications. Artificial Intelligence in Medicine, 145, 102644.
17. Wang, F., Kaushal, R., & Khullar, D. (2022). Should health care demand interpretable artificial intelligence or accept black box medicine? Annals of Internal Medicine, 175(4), 587–589.
18. Inala, R., Kaulwar, P. K., Nagabhyru, K. C., Adusupalli, B., & Arun Raj, S. R. (2025, October). Leveraging IEC 61850 for Interoperable and Resilient Smart Grid Communication Architecture. In International Conference on Microelectronics, Electromagnetics and Telecommunication (pp. 549-566). Cham: Springer Nature Switzerland.
19. Beam, A. L., & Kohane, I. S. (2022). Translating artificial intelligence into clinical care. JAMA, 328(14), 1395–1396.
20. Rajkomar, A., Dean, J., & Kohane, I. (2022). Machine learning in medicine. New England Journal of Medicine, 386(14), 1347–1358.
21. Nabende, P., & Wanyama, T. (2008). An expert system for diagnosing heavy-duty diesel engine faults. In Advances in computer and information sciences and engineering (pp. 384-389). Dordrecht: Springer Netherlands.

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

22. Topol, E. (2023). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 29(1), 44–56.
23. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2021). Deep learning for healthcare: Review, opportunities and challenges. Briefings in Bioinformatics, 22(6), bbab048.
24. Johnson, A. E. W., Pollard, T. J., & Mark, R. G. (2021). Reproducibility in critical care machine learning using electronic health records. NPJ Digital Medicine, 4(1), 103.
25. Amistapuram, K., Pandiri, L., Raju, V. R., Paleti, S., Singireddy, S., & Sheelam, G. K. (2025). AI-Based Cloud Infrastructure and MLOps Frameworks for Scalable Data Engineering Across Banking and Insurance. In 2025 IEEE International Conference on Communication Networks and Computing (CNC) (pp. 186–192). IEEE. 2025 IEEE International Conference on Communication Networks and Computing (CNC). https://doi.org/10.1109/cnc68716.2025.11484532
26. Xiao, C., Choi, E., & Sun, J. (2021). Opportunities and challenges in developing deep learning models using electronic health records. Journal of the American Medical Informatics Association, 28(1), 95–102.
27. Sendak, M. P., Gao, M., Nichols, M., Lin, A., & Balu, S. (2022). Machine learning in health care: A critical appraisal of challenges and opportunities. eGEMs, 10(1), 1–12.
28. Sudhakar, A. V. V., Inala, R., Verma, A. K., Nag, K., Pandey, V., & Anand, P. S. (2025, September). Hybrid Rule-Based and Machine Learning Framework for Embedding Anti-Discrimination Law in Automated Decision Systems. In 2025 International Conference on Intelligent Communication Networks and Computational Techniques (ICICNCT) (pp. 1-6). IEEE.
29. Goldstein, B. A., Navar, A. M., Carter, R. E., & Moving Beyond Regression Techniques Working Group. (2022). Opportunities and challenges in developing risk prediction models using electronic health record data. Journal of the American Heart Association, 11(5), e024150.
30. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2021). Deep learning in electronic health records: A systematic review of methods, applications, and challenges. Journal of Biomedical Informatics, 118, 103799.
31. Aguirre, R. R., & Starren, J. (2021). Electronic health records and precision medicine: Current applications and future opportunities. Yearbook of Medical Informatics, 30(1), 170–179.
32. Kolla, S. H., & Mangala, N. (2025). DESIGNING AUTONOMOUS LLM AGENT FRAMEWORKS USING GEN AI PIPELINES TO ENHANCE CUSTOMER SERVICE MANAGEMENT AND KNOWLEDGE WORKFLOWS. Lex Localis-Journal of Local Self-Government, 23, 9719-9733.
33. Jensen, P. B., Jensen, L. J., & Brunak, S. (2021). Mining electronic health records: Towards better research applications and clinical care. Nature Reviews Genetics, 22(6), 395–410.
34. Gottesman, O., Kuivaniemi, H., Tromp, G., Faucett, W. A., Li, R., Manolio, T. A., Sanderson, S. C., Kannry, J., Zinberg, R., Basford, M. A., & Bottinger, E. P. (2021). The Electronic Medical Records and Genomics (eMERGE) Network: Past, present, and future. Genetics in Medicine, 23(2), 234–242.
35. Wang, L., Porter, B., Maynard, C., Evans, G., Bryson, C., Sun, H., Gupta, I., & Balu, S. (2022). Predictive analytics using electronic health records for clinical decision support: A systematic review. Journal of Medical Systems, 46(4), 28.
36. Kumar, I., Nagabhyru, K. C., IG, N., MV, P., & KV, S. (2025, October). Adaptive Meta-Knowledge Transfer Network with Feature Hallucination and Attention for Low-Shot Object Detection in Aerial Images. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1-6). IEEE.
37. Zhang, Y., Chen, R., & Xu, H. (2022). Artificial intelligence and electronic health records in

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

precision medicine: Current trends and future directions. Artificial Intelligence in Medicine, 126, 102271.
38. Kaissis, G., Makowski, M., Rückert, D., & Braren, R. (2022). Secure, privacy-preserving and federated machine learning in medical imaging and electronic health records. Nature Machine Intelligence, 4(2), 119–130
39. Loganathan, R. (2025). AGENTIC AI FRAMEWORKS FOR AUTONOMOUS RISK DETECTION AND COMPLIANCE REMEDIATION IN ENTERPRISE DATA CENTER OPERATIONS. Lex Localis-Journal of Local Self-Government, 23 (S6), 9672–9697.
40. Wornow, M., Xu, Y., Thapa, R., Patel, B., Steinberg, E., Fleming, S., & Shah, N. H. (2023). The shaky foundations of clinical foundation models: A survey of large language models in healthcare. NPJ Digital Medicine, 6(1), 201.
41. Luo, Y., Xin, Y., Joshi, R., Celi, L. A., & Szolovits, P. (2023). Predictive modeling using electronic health record data: Challenges and opportunities. Journal of the American Medical Informatics Association, 30(5), 903–914.
42. Amistapuram[1], K., Kolla, T., Bandi, V. D. V. K., Kolla[4], S. K., & Rani, P. S. Journal of Rare Cardiovascular Diseases.
43. Li, X., Ding, Y., Shao, Y., & Wu, C. (2023). Intelligent electronic health record analytics using deep learning for personalized healthcare. IEEE Access, 11, 51432–51447.
44. Ahmad, M. A., Eckert, C., & Teredesai, A. (2023). Explainable artificial intelligence for healthcare: A review of applications using electronic health records. ACM Computing Surveys, 56(2), 1–37.
45. Davuluri, P. N. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems.
46. Nguyen, P., Tran, T., Wickramasinghe, N., & Venkatesh, S. (2023). Deep representation learning from electronic health records for predictive healthcare. Artificial Intelligence in Medicine, 138, 102500.
47. Chen, I. Y., Joshi, S., Ghassemi, M., & Goldenberg, A. (2023). Treating health disparities with artificial intelligence. Nature Medicine, 29(3), 547–556.
48. Kallianos, K., Mongan, J., Antani, S., Henry, T., Taylor, A., Abuya, J., & Kohli, M. (2023). How far have we come? Artificial intelligence for healthcare using electronic health records. Radiology: Artificial Intelligence, 5(2), e220145.
49. Garapati, R. S., & Kanna, S. R. A Digital Twin-Enabled Predictive Maintenance Framework Leveraging Multi-Agent Reinforcement Learning and Industrial IoT Data.
50. Xu, J., Wang, Y., Zhao, H., & Li, Q. (2024). Multimodal learning from electronic health records for precision clinical prediction. IEEE Journal of Biomedical and Health Informatics, 28(3), 1521–1533.
51. Chen, T., Liu, Y., Wang, Z., & Zhang, H. (2024). Intelligent health analytics using large language models and electronic health records. Journal of Biomedical Informatics, 149, 104596.
52. Huang, K., Altosaar, J., & Ranganath, R. (2024). Clinical representation learning from longitudinal electronic health records. Nature Communications, 15(1), 1084.
53. Kolla, T. (2025). Generative AI for Intelligent Medical Coding and Healthcare Analytics. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 8(6), 13285-13299.
54. Islam, M. M., Poly, T. N., Walther, B. A., Li, Y. J., & Li, Y. C. (2024). Artificial intelligence in electronic health records for disease prediction and precision medicine: A systematic review. Healthcare, 12(4), 421.
55. Yu, K. H., Beam, A. L., & Kohane, I. S. (2024). Artificial intelligence in healthcare: Past, present, and future. Nature Biomedical Engineering, 8(1), 18–31.

P-ISSN: 3050-9572
E-ISSN: 3050-9580
EUROPEAN DATA SCIENCE JOURNAL
Volume: 03 Issue: 01
RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

56. Reddy, V. A. R., & Kolla, S. K. (1984). Infrastructure-As-Code Practices For Regulated Healthcare Cloud Environments. Metallurgical and Materials Engineering, 30 (4), 1028–1042.
57. Patel, B. N., Rosenberg, L., Willcox, G., Baltaxe, D., Lyons, M., Irvin, J., Rajpurkar, P., Amrhein, T., et al. (2024). Human–artificial intelligence collaboration in precision healthcare and clinical decision support. The Lancet Digital Health, 6(2), e95–e106.
58. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G. S., Thrun, S., & Dean, J. (2021). A guide to deep learning in healthcare. Nature Medicine, 27(5), 749–760.
59. Ashokkumar, S., & Amistapuram, K. (2025, October). Attention-Guided Spatial Temporal Framework for Deepfake Detection on Social Video Platforms. In 2025 International Conference on Communication, Computer, and Information Technology (IC3IT) (pp. 1-6). IEEE.
60. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2021). Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology, 6(2), 230–243.
61. Liu, X., Rivera, S. C., Faes, L., & Denniston, A. K. (2021). Reporting guidelines for clinical artificial intelligence studies. Nature Medicine, 27(8), 1322–1324.
62. Nigam, N., Sireesha, B., Ediga, P., Segireddy, A. R., & Bokde, S. (2025, December). Comparative Evaluation of Cloud Security Algorithms Using Multiple Classifiers with an Optimized Intrusion Detection System. In 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG) (pp. 1-6). IEEE.
63. Chen, M., Decary, M., & Wang, L. (2022). Artificial intelligence in healthcare: Challenges and opportunities for precision medicine. Frontiers in Digital Health, 4, 868091.
64. Ribeiro, M. T., Singh, S., & Guestrin, C. (2022). Explainable artificial intelligence for trustworthy clinical decision support systems. AI Magazine, 43(3), 281–294.
65. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2022). Transparent reporting of multivariable prediction models in healthcare. BMJ, 378, e067981.
66. Mangalampalli, B. M., & Kolla, S. K. (2025). Large Language Models for Automated Healthcare Data Dictionary Generation and Maintenance. Vascular and Endovascular Review, 8(20s), 363-375.
67. Topol, E. J. (2022). Deep medicine and the future of precision healthcare. Nature Reviews Clinical Oncology, 19(10), 613–614.
68. Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., & Goldenberg, A. (2022). Do no harm: A roadmap for responsible machine learning in healthcare. Nature Medicine, 28(10), 1990–2001.
69. Singh, H., Bose, D., Nagubandi, A. R., Prabhu, S., & Naik, S. G. Cryptocurrency Market Spillovers: Risk Contagion Across Global Financial Systems.
70. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2023). AI in health and medicine. Nature Medicine, 29(1), 31–38.
71. Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2023). The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, 5(11), e745–e750.
72. Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616(7956), 259–265.
73. Reddy, V. A. R. (2025). Journal of Rare Cardiovascular Diseases. Health, 5(3), 402-422.
74. Bohr, A., & Memarzadeh, K. (2023). Artificial intelligence in precision medicine: Applications and challenges. Artificial Intelligence in Healthcare, 2nd ed., Academic Press, 341–365.

P-ISSN: 3050-9572 E-ISSN: 3050-9580

EUROPEAN DATA SCIENCE JOURNAL

Volume: 03 Issue: 01

RECEIVED: JANUARY 25 REVISED: FEBRUARY 07 ACCEPTED: FEBRUARY 26 PUBLISHED: MARCH 12

75. Krittanawong, C., Virk, H. U. H., Bangalore, S., Wang, Z., Johnson, K. W., Pinotti, R., Zhang, H., Kaplin, S., Narasimhan, B., & Tang, W. H. W. (2023). Machine learning prediction in cardiovascular medicine using electronic health records. European Heart Journal – Digital Health, 4(2), 101–113.
76. LEBCIR, I., Shah, C. A., & Appa Rao Nagubandi, D. S. M. D. FinTech and Financial Inclusion: Empirical Evidence from Emerging Markets.
77. Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F. K., & Mahmood, F. (2024). Synthetic data in machine learning for medicine and healthcare. Nature Biomedical Engineering, 8(2), 111–123.
78. Luo, J., Wang, F., & Sun, J. (2024). Artificial intelligence for healthcare data analytics: Advances and future directions. ACM Computing Surveys, 56(8), 1–36.
79. Yang, Q., Steinfeld, A., Zimmerman, J., & Rosé, C. P. (2024). Re-examining human-AI collaboration in clinical decision-making. Communications of the ACM, 67(3), 54–63.
80. Bandi, V. D. V. K. AI-Based Anomaly Detection Frameworks in Distributed Enterprise Data Systems.
81. Ge, J., Zhang, Y., Xu, H., & Wang, F. (2024). Large language models for electronic health record intelligence: Opportunities and challenges. Journal of Biomedical Informatics, 150, 104622.
82. Lee, P., Bubeck, S., & Petro, J. (2024). Benefits, limits, and risks of GPT-4 as an AI assistant for medicine. New England Journal of Medicine, 390(13), 1233–1239.
83. GARAPATI, R. S. SYNERGETIC INTELLIGENCE Converging AI, Cloud, IoT, and Smart Automation for Real-Time Futures. CANEDA GLOBAL JOURNAL GROUP.
84. Wang, Z., Li, X., Chen, H., & Sun, J. (2025). Intelligent electronic health record engineering for precision health analytics: Recent advances and future perspectives. Journal of Biomedical Informatics, 156, 104731.
85. Zhang, H., Wu, J., Liu, Y., & Wang, F. (2025). Artificial intelligence–driven precision health analytics using multimodal electronic health records. IEEE Journal of Biomedical and Health Informatics, 29(2), 945–958.