



Methods and Challenges in Evaluating Electronic Health Record Systems: A Systematic Literature Review

Ade Dita Puteri^{1*}, Lira Mufti Azzahri Isnaeni¹

¹Universitas Pahlawan Tuanku Tambusai, Kampar, Indonesia.

Received: August 27, 2025

Revised: December 13, 2025

Accepted: January 16, 2026

Published: January 16, 2026

Corresponding Author:

Ade Dita Puteri

adedita10@gmail.com

DOI: [10.29303/jppipa.v11i12.12673](https://doi.org/10.29303/jppipa.v11i12.12673)

© 2025 The Authors. This open access article is distributed under a (CC-BY License)



Abstract: Systems due to their ability to increase the accessibility of clinical information, support data-driven decision-making, and potentially improve patient care quality and safety. However, EHR system evaluation remains highly complex due to the dynamic interaction between technical, organizational, and human factors. This article presents a systematic literature review aimed at identifying, classifying, and synthesizing EHR evaluation methods used in various healthcare contexts, while also uncovering key challenges in their implementation and assessment processes. Following PRISMA guidelines, this study examined scholarly articles published between 2013 and 2023 and indexed in reputable academic databases. The review reveals a wide variety of EHR evaluation approaches, including usability testing, system performance measurement, cost-benefit analysis, and qualitative methods such as user satisfaction surveys, in-depth interviews, and ethnographic studies. However, various challenges remain, including limited interoperability, data privacy and security issues, deficiencies in methodological rigor, clinical user resistance, and ethical dilemmas in health information management. Furthermore, the lack of a standardized evaluation framework further complicates comprehensive and ongoing assessment efforts. This study emphasizes the need to develop an integrated, multidisciplinary, and ethically sound EHR evaluation model to ensure optimal system benefits while minimizing future implementation risks.

Keywords: Electronic health records; Evaluation methods; Health informatics; Systematic literature review

Introduction

The development of information technology in the health sector has brought a significant transformation in patient data management, one of which is through the implementation of Electronic Health Records (EHR). EHR is a digital system designed to integrate, store, and manage patients' medical records electronically, thereby facilitating access, efficiency, and accuracy of health information (Reegu et al., 2023; Shaikh et al., 2022). Unlike traditional paper-based medical records, EHRs allow for real-time data exchange between healthcare facilities, speed up clinical decision-making processes, and improve coordination between healthcare workers. Thus, EHR serves as a modern solution to improve the quality of healthcare services while reducing the risk of

medical errors (Kruse et al., 2016). In addition, the use of EHR also provides strategic value in aspects of evidence-based medicine, big data analysis, and improving the security of patient health information (Sezgin & Özkan, 2013). EHRs are capable of storing patient longitudinal data that can be used for research, chronic disease monitoring, and evaluation of treatment effectiveness. On the other hand, this technology supports interoperability between different health systems, thus enabling the integration of information on a national and global scale. However, EHR implementation still faces a number of challenges, such as high costs, medical personnel resistance, and patient data privacy and security issues that need to be addressed through strict policies and standards (Shankar et al., 2025; Cahill et al., 2025).

How to Cite:

Puteri, A. D., & Isnaeni, L. M. A. (2025). Methods and Challenges in Evaluating Electronic Health Record Systems: A Systematic Literature Review. *Jurnal Penelitian Pendidikan IPA*, 11(12), 140-157. <https://doi.org/10.29303/jppipa.v11i12.12673>

Electronic Health Records (EHRs) play an important role in improving the efficiency of healthcare services by accelerating access to patient information, reducing duplication of examinations, and reducing operational costs (Kruse et al., 2016). EHR systems allow medical personnel to obtain patient health histories quickly and accurately, thereby speeding up the clinical decision-making process (Meystre et al., 2008; Koleck et al., 2019). Research shows that the implementation of EHR significantly reduces the time required for the management of medical administration and documentation, so that healthcare workers can focus more on patient care. Thus, the resulting efficiency not only impacts the service provider, but also improves patient satisfaction. In addition to efficiency, EHRs also contribute to improved diagnostic accuracy and coordination between healthcare providers (Goldstein et al., 2017). EHRs support the integration of clinical data from various sources, such as laboratory, radiology, and consultation records, providing a comprehensive picture of a patient's condition. This helps doctors reduce the risk of medical errors due to limited information and improve the quality of clinical decision-making. Furthermore, EHR allows data sharing between healthcare facilities, supports referral systems, and improves care coordination, especially for patients with chronic diseases that require multidisciplinary management (Goldstein et al., 2017; Kruse et al., 2016). With this capability, EHRs are seen as a key instrument in realizing safer, more effective, and more sustainable healthcare services (Lewis et al., 2023; Cook et al., 2022).

Evaluation of an Electronic Health Records (EHR) system is a complex process because it includes a variety of interrelated technical, managerial, and clinical aspects. From the technical side, the evaluation involves interoperability, security, and usability factors (Aliabadi et al., 2020; Payrovnaziri et al., 2020). Interoperability is crucial because EHRs must be able to communicate with a variety of different health information systems, both within and across institutions. Data security is also a major concern considering that EHRs store highly sensitive patient information, making them vulnerable to the threat of leaks and cyberattacks. In addition, the usability aspect is directly related to user acceptance, where unintuitive interface design can reduce the effectiveness of system utilization by medical personnel. From a managerial and clinical perspective, EHR evaluations must consider the cost of implementation, the availability of training, and their impact on the quality of service. EHR implementation requires a major investment not only in software and hardware, but also in healthcare workforce training to ensure optimal adoption (Mayer et al., 2020). Limited human resources and resistance to change are often obstacles to the full utilization of this system. Clinically, evaluations should

measure the extent to which EHRs are able to improve the quality of service, including the reduction of medical errors, improved care coordination, and patient satisfaction. It is this complexity that demands a multidimensional evaluation approach, taking into account technical, managerial, and clinical aspects (Torab-Miandoab et al., 2025; Reegu et al., 2023; Dewanto et al., 2024; Elfira & Santosa, 2023; Ichsan et al., 2023; Zulyusri et al., 2023).

The implementation and evaluation of Electronic Health Records (EHR) in various countries face several challenges that cannot be ignored. One of the main obstacles is the limitation of evaluation methodologies, where there is no universal standard that can be used to comprehensively assess the effectiveness, efficiency, and clinical impact of EHRs (Haque et al., 2022). Many studies use different approaches, ranging from qualitative analysis based on user perception to quantitative studies that assess indicators of service efficiency, making results are often difficult to compare consistently. This results in a lack of clarity in determining the indicators of successful EHR implementation, both from the technical, managerial, and clinical sides. In addition to the methodological aspect, other challenges that arise are ethical issues and resistance from health workers. EHRs store large amounts of patient data, raising concerns about confidentiality, privacy, and the potential for misuse of health information (Tapuria et al., 2021; Ettaloui et al., 2024). This data security issue is increasingly important amid growing cyber threats to health information systems. On the other hand, healthcare worker resistance often arises due to changes in workflows, increased administrative burden, and lack of adequate training. Many medical professionals feel that EHRs reduce interaction time with patients as the focus shifts to digital data entry. Therefore, these challenges must be addressed through policy formulation, more user-friendly system design, and effective change management support (Zarour et al., 2021).

A number of previous studies have shown that the evaluation of Electronic Health Records (EHR) is carried out with a variety of methodological approaches that produce different findings. For example, some research emphasizes that although EHRs are able to improve access to information and service efficiency, the evaluation methods used are often limited to administrative measures such as time and cost efficiency, so they do not fully capture clinical impact or patient satisfaction (Abramson et al., 2012). Meanwhile, the challenge of EHR evaluation is not only in the technical aspect, but also in how the system affects organizational changes and daily medical practice. This suggests that previous studies have tended to focus on specific dimensions, without providing a

comprehensive picture of the effectiveness of EHRs. Additionally, several studies highlight practical challenges in the implementation and evaluation of EHRs. Adler-Milstein & Jha (2017) found that the issue of interoperability and data security is still the main obstacle that hinders the full utilization of EHRs in various countries (V.H. et al., 2010). Kruse et al. (2016) through a systematic review confirmed that medical personnel resistance, lack of training, and ethical issues related to patient privacy are factors that consistently arise in various contexts of EHR implementation. Thus, previous research provides an important basis indicating that there are still gaps in the literature, particularly in bringing together diverse evaluation methods and identifying key challenges in a more structured manner. Based on this research, this study aims to identify Methods and Challenges in Evaluating Electronic Health Record Systems: A Systematic Literature Review.

Method

This systematic literature review synthesizes existing approaches for evaluating electronic health record (EHR) systems and examines the most frequently cited challenges encountered during their rollout. Following the PRISMA framework described by Page in 2020, the study aimed to maintain transparency and methodological rigour at every stage of the review (see Figure 1).

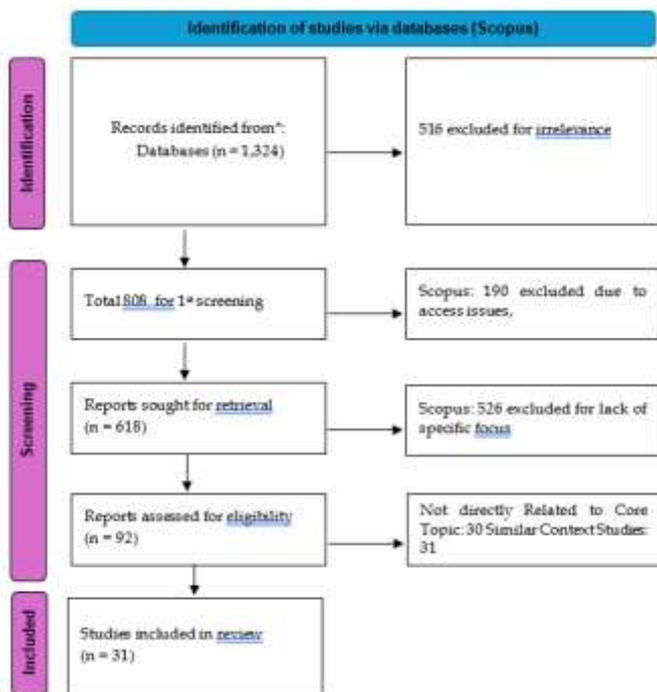


Figure 1. PRISMA process

An extensive literature review was conducted, targeting the Scopus database with a clearly defined keyword set. The retrieval string, articulated with the terms electronic, medical, records, database, and evaluation joined by the Boolean operator AND, was designed to capture relevant material with minimal off-topic noise. This initial query returned 1,324 items from the database. In the interest of reproducibility and transparency, only open-access publications were kept, trimming the count to 808 items. Further refinement targeted original research articles, excluding editorials, conference abstracts, book chapters, and opinion pieces, which brought the total to 618 works.

The eligibility of these 618 items was assessed through a two-step review procedure carried out independently by two researchers. During the first stage, titles and abstracts were screened to determine whether they examined the evaluation, validation, or general assessment of electronic health record systems from technical, organisational, or user-centred angles. Studies focused solely on disease-specific workflows that did not analyse system-level features were set aside. At the end of the first pass, 92 articles advanced to full-text review for the final decision.

In the second stage of the review, two independent researchers examined each paper in its entirety. Whenever their interpretations diverged, they met to discuss the differences until they agreed on a single reading. To move forward, a study had to document either the methods it used to assess electronic health record (EHR) systems or the obstacles encountered during their rollout and everyday use. Once these criteria were applied, 32 papers advanced to the final analysis.

Throughout the screening and selection process, the flow of study identification, screening, eligibility, and inclusion followed the structure recommended by PRISMA. The journey began with 1,324 records; 808 of those turned out to be openly available. After filtering out duplicates and focusing on original research, 618 articles remained. Review of titles and abstracts yielded 92 candidates, and a final check of the full texts confirmed 31 studies for inclusion in this review.

Data extraction then proceeded in a consistent manner, guided by a template created before the review. Collectors recorded authors, publication year, purpose, methods, evaluation criteria, challenges reported, and acknowledged limitations. They paid particular attention to matching each entry to the specific research questions, namely how EHR systems have been evaluated (RQ1) and what hurdles their adopters have met (RQ2). Finally, thematic analysis brought the individual data points together into a coherent overview.

Result and Discussion

Result

To outline what is currently known about the assessment, design, and persistent hurdles of electronic health record (EHR) systems, the present review examines a curated set of thirty-one empirical studies drawn from diverse research traditions and patient-care environments. The entries listed in Table 1 include rich qualitative appraisals such as Xanthidou et al. (2018), practical design work coupled with usability trials as in Jokar et al. (2023) and Allorto & Wise (2015), machine-learning models aimed at clinical prediction in Xiang et al. (2019) and Zang et al. (2024), designs anchored in blockchain technology by Regueiro et al. (2021), ambitious projects seeking large-scale data harmonisation by Muñozerro-Muñiz et al. (2020) and Ferrara et al. (2023). Addressing the subtler side of knowledge extraction, several articles report advances in natural-language processing as Zhang et al. (2022) and Siefriidt et al. (2020), whereas others validate routine administrative records by comparing them to clinical gold standards as shown by Schwartz et al. (2019) and Dart et al. (2023). To situate these findings within an evolving landscape, the sample also incorporates systematic literature reviews.

To further characterize how electronic health record (EHR) systems and databases have been evaluated in the existing literature, this part (Table 1) systematically maps the methodological approaches employed across the 31 included studies. Table 1 provides a detailed summary of the types of evaluation methods, decision-making approaches, and unique methodological characteristics that distinguish each work. The studies span a wide range of methodologies, including qualitative interviews grounded in phenomenological inquiry as demonstrated by Xanthidou et al. (2018), applied system development and usability testing frameworks as written by Jokar et al. (2023) and Allorto & Wise (2015), machine learning model adaptation and validation studies developed by Xiang et al. (2019) and Zang et al. (2024), and database infrastructure benchmarking conducted by Freire et al. (2016) and Muñozerro-Muñiz et al. (2020). Other contributions reflect the growth of emerging fields such as blockchain system evaluation explored by Regueiro et al. (2021) and heterogeneous graph-based EMR modeling developed by Cho et al. (2022). Meanwhile, systematic reviews by Wang & Zhang (2024) and Alzoubi et al. (2019), along with expert consensus guidelines proposed by Deignan et al. (2019), provide critical thematic overviews and policy guidance.

After analyzing Table 1, there are some evaluation methods identified, which are discussed below. To better understand how researchers have assessed

electronic health record (EHR) systems and the databases that underlie them, this section maps the methods used in the thirty-one studies included in Table 1. The review reveals a rich methodological Xanthidou et al. (2018) carried out qualitative interviews anchored in phenomenological principles, capturing the nuances of how users interpret their everyday encounters with the records. Jokar et al. (2023) and later Allorto & Wise (2015) applied usability-testing frameworks to propose modular design, embedding structured database components within intensive-care-unit workflows. Xiang et al. (2019) and Zang et al. (2024) shifted the focus to machine learning, adapting and validating predictive algorithms trained on real-world EHR data. Freire et al. (2016) and Muñozerro-Muñiz et al. (2020) then performed benchmarking exercises, measuring throughput and latency across competing database architectures. Regueiro et al. (2021) examined blockchain-integrated electronic-medical-record platforms, while Cho et al. (2022) introduced heterogeneous graph modelling to represent multifaceted patient profiles. Systematic reviews by Wang & Zhang (2024), together with Alzoubi et al. (2019), synthesised thematic insights from disparate works, and Deignan et al. (2019) extended that overview by highlighting emerging gaps and future directions. Expert working groups have put forward consensus guidelines intended to shape electronic health record policy and the practical steps taken during system roll-out.

The use of NLP techniques with EHRs has made it possible to systematically extract clinical knowledge from unstructured text. To improve semantic interoperability, Zhang et al. (2022) applied BERT as well as BiLSTM-CRF models to convert clinical narratives into knowledge representation. Siefriidt et al. (2020) validated multi-terminological concept extraction tools and showed that automatic coding of French family medicine consultations is possible. In sequential clinical event mining, Bonomi & Jiang (2018) proposed a graph-based method to derive significant patient trajectories within hospital data. Also, Kehl et al. (2024) demonstrated a privacy-preserving model distillation framework for drawing out cancer outcomes from clinical notes, thereby underscoring secure AI application possibilities in healthcare. Machine Learning for Prediction, Phenotyping, and Modeling Researchers have turned to machine learning to refine patient outcome forecasts, disease classification, and long-term EHR modelling. Xiang et al. (2019) sharpened time-sensitive clinical embeddings, boosting the accuracy of both classification and forward-looking tasks. Concurrently, Cho et al. (2022) built heterogeneous graphs to anticipate major cardiovascular events. Looking at therapy changes, Cui et al. (2023) introduced

SimPA, which supplements a patient's timeline with records from clinically similar individuals, yielding more accurate transition forecasts. In an extensive overview, Alzoubi et al. (2019) and associates confirmed that hybrid phenotyping systems, jointly analyzing structured fields and free-text notes, still outpace purely

single-source pipelines. Lastly, Zang et al. (2024) examined how well models transfer between hospitals, concluding that regularised logistic regression often generalises farther than many advanced deep-learning architectures (Alqahtani et al., 2017; Katurura & Cilliers, 2018).

Table 1. Methods of evaluation methods

Authors (Year)	Methodology Used	Type of Decision-Making Approach	Notes/Special Characteristics
Xanthidou et al. (2018)	Phenomenological Qualitative Study (Semi-structured Interviews)	Professional evaluation based on practitioners' experiences	Emphasized open-ended, experiential data collection; validated by multi-country academic reviewers; included in-person and email-based interviews
Zhang et al. (2022)	Knowledge Organization Method + NLP (BiLSTM-CRF, BERT)	Structured knowledge extraction and coding; database construction	Focused on standardizing natural language clinical descriptions into structured, codable knowledge using machine learning and terminology harmonization
Jokar et al. (2023)	Applied Development Study + Usability Evaluation (USE Questionnaire)	Functional and non-functional requirements analysis; system implementation and evaluation	Focused on system development for specific EMR module (anesthesia), using clinical and technical expert validation, and end-user feedback
Xiang et al. (2019)	Adapted word2vec (Skip-gram), PPMI-SVD, FastText, all time-sensitive	Clinical concept embedding training; machine learning model extension	Modified traditional ML models to incorporate temporal dependency; evaluated embeddings via clustering-based, classification-based, and predictive modeling-based tasks
Regueiro et al. (2021)	Blockchain Architecture Development + Prototype Testing (Ethereum, Smart Contracts, Synchronizer Mechanism)	Infrastructure evaluation under network failure conditions	Focus on real-time synchronization across blockchain and local EMR databases; used Proof of Authority consensus; evaluated system performance on latency, throughput, and user satisfaction
Muñoyerro-Muñiz et al. (2020)	Deterministic Data Linkage + Data Warehouse Architecture + Business Intelligence Reporting	Health data integration and analysis infrastructure	Developed three-layer architecture (data, application, presentation); used Oracle 11g DB, MicroStrategy BI; ensured data governance and pseudonymization for patient privacy
Cho et al. (2022)	Heterogeneous Graph Construction + HinSAGE (Heterogeneous GraphSAGE)	Graph-based EMR integration; supervised link prediction task	Constructed patient-centered bipartite graph database; HinSAGE model trained to predict cardiovascular outcomes based on node and edge attributes; graph visualization and querying via Neo4j
Allorto & Wise (2015)	Relational Database Design (FileMaker Pro®) + Clinical Workflow Integration + Field Audit Evaluation	Systematic relational design aligned with ICU clinical workflow	Designed database tables and fields to mirror patient journey (referral, admission, discharge); training of clinical staff; field audit to assess data completeness and compliance over two system versions
ASGE Technology Committee et al. (2016)	Systematic Literature Review + Technology Status Evaluation	Benchmarking evaluation across EEMR vendors and functionalities	Focused on hardware/software integration, image capture, HL7/EHR interoperability, quality measure tracking, practice management modules
Burt et al. (2017)	Planned Systematic Review using PRISMA-P Framework	Systematic mapping of usability assessment techniques for EMR visualization tools	Focused on capturing usability evaluation tools based on criteria like efficiency, effectiveness, satisfaction; cross-comparing visualisation categories (static, semi-static, dynamic)
Pacurariu et al. (2018)	Systematic Database Review and Descriptive Analysis	Structural evaluation of EHDs' regulatory suitability	Focused on regulatory relevance (availability, accessibility, quality validation, potential for CDM harmonization); data cross-verified with database owners; involved ENCePP network endorsement
Bonomi & Jiang (2018)	Graph-based Pattern Matching + Null Hypothesis Statistical Test	Data mining and statistical evaluation of EHR sequences	Constructed multi-partite graph for efficient pattern search; statistical ranking of patient matches using p-values; evaluated using real hospital data (MIMIC-III); focused on retrieving clinically significant patient

Siefridt et al. (2020)	Natural Language Processing (MTCE) + Manual Annotation Validation	Concept extraction and quality assessment using multiple terminologies	Evaluated 15,000 annotations using a web-based tool; assessed inter- and intra-terminological contributions; dual-evaluator validation to ensure consistency; focused on French-language EMR data sequences
Schwartz et al. (2019)	Validation Study (Diagnostic Accuracy Metrics) + Trend Analysis (Poisson Regression)	Comparative analysis of prescribing vs dispensing databases using diagnostic test evaluation methods	Focused on assessing EMR reliability for studying antibiotic use; gold standard ODB dispensing data used; stratified analysis by physician prescribing volume; applied Poisson models to identify trend discrepancies
Nakamura et al. (2015)	Quantitative Comparative Analysis of Influenza Surveillance Systems	Comparative validation of national influenza surveillance accuracy	Evaluated three datasets (NOSSID, PS, NDBEMC); adjusted for coverage differences; applied statistical ratio comparisons; addressed time lag and selection bias in sentinel surveillance
Wang & Zhang (2024)	Systematic Literature Review + Task-wise Analysis	Comprehensive mapping of LLM healthcare applications	Categorized applications into medical QA, dialog summarization, EHR generation, imaging, clinical reasoning, scientific research, education; assessed models (e.g., ChatGPT, GPT-4, MedPaLM) against tasks and datasets
Deignan et al. (2019)	Expert Consensus-Based Guideline Formulation	Best practice recommendations for genomic data reevaluation and reanalysis	Developed by ACMG Laboratory Quality Assurance Committee; addressed processes at variant-level and case-level; included recommendations on reporting, billing, and periodic policy review
Ammann et al. (2018)	Chart Validation Study + Positive Predictive Value (PPV) Statistical Evaluation	Diagnostic code validation within a healthcare administrative database	Focused on venous thromboembolism (VTE) events; evaluated how well ICD-9-CM administrative codes predicted real clinical events based on manual chart review; stratified by diagnosis code position
Huber et al. (2022)	Field Survey + Descriptive Quality Assessment + Data Flow Mapping	Evaluation of routine data readiness for impact evaluation	Comprehensive register and database inventory across 24 clinics; assessed by direct site observation and documentation review; developed ideal data flow model for chronic disease and HIV patient management
Sander & Wauer (2019)	System Development (Ontology-SQL) + Semantic Query Evaluation + Annotation Quality Testing	Integrating ontology-based semantic search directly into relational databases via SQL extension	Created Ontology-SQL (O-SQL) parser that rewrites semantic queries to standard SQL; used German medical dataset for system validation; assessed by sensitivity, specificity, F-score; semantic expansion through ontology integration
Parejo et al. (2020)	Tag Cloud-Based Interface Development + Retrieval Evaluation Metrics + User Survey	Graphical exploration tool for EMR using multi-term tag clouds generated automatically	Developed based on WAPO-Structure; incorporated semantic preprocessing (synonyms, acronyms); validated by precision, recall, F1-score, and 23-user Likert-scale survey; novel multi-term rather than mono-term tag clouds
Cui et al. (2023)	Similar Patient Augmentation with RNN + Transformer Backbone	Temporal sequence modeling and semantic augmentation with graph attention	Developed SimPA architecture to incorporate similar patients (same doctor, same treatment line) using Transformer; dynamic update via GRU for patient's visit sequence; supervised by cross-entropy loss
Freire et al. (2016)	Experimental Database Benchmarking (Centralized and Distributed Testing)	Performance evaluation of NoSQL databases for archetype-based EMR data storage and querying	Tested BaseX, eXistdb, BerkeleyDB XML (XML-based NoSQL) and Couchbase (JSON-based NoSQL) against MySQL; measured response times across datasets from 10k to 4.2 million records; included cluster scaling analysis for Couchbase
Kim et al. (2021)	ETL Process to OMOP CDM + Predictive Modeling (ML) Validation	Conversion and validation of PSG data into a standardized multi-center research format	Developed ETL for 11,797 PSGs; manually mapped standard and custom concepts; validated usefulness via Random Forest, GBM, Lasso models; utilized OHDSI PatientLevelPrediction package; predictive outcome was cardio-neuro-metabolic disease
Nash et al. (2017)	Pragmatic Cluster-Randomized Controlled Trial + Embedded Qualitative Process Evaluation	Real-world trial with minimal interference to practice patterns	Randomized 34 clinics using minimization; embedded clinical decision support (CDS) tools, audit and feedback (SAFIRE reports), patient handouts; process evaluation via semi-structured physician interviews based on Normalization Process Theory

Nash et al. (2017)	Pragmatic Cluster-RCT + Mixed-Methods Process Evaluation (Normalization Process Theory)	Evaluation of EMR-based multifaceted intervention to improve CKD care	Developed CKD toolkit with EMR-based decision support, SAFIRE audit-feedback reports, patient handouts, and physician implementation support; outcome analysis using generalized estimating equations with cluster adjustments; embedded qualitative interviews to assess toolkit uptake barriers and facilitators
Alzoubi et al. (2019)	Systematic Literature Review + Structured Synthesis	Synthesis of phenotyping system development based on structured, unstructured data, rule-based and machine learning methods	Categorized studies by feature extraction (structured, BoW, keyword search, concept extraction) and classification techniques (rule-based, supervised ML, unsupervised ML, hybrid); evaluated validation approaches (gold standards, cross-validation)
Kehl et al. (2024)	Teacher-Student Model Distillation + Supervised Learning with NLP Transformers	Secure training of clinical AI models without direct exposure to PHI	Teacher models trained on PHI-labeled data (DFCI); public dataset (MIMIC-IV) labeled by teacher; student models trained on public labels; student models evaluated externally on MSK data; tested imaging reports and oncologist notes separately; also validated survival prediction using c-index
Dart et al. (2023)	Validation Study: Cross-Sectional Design + Algorithm Evaluation + Statistical Diagnostic Accuracy Analysis	Development and testing of 18 administrative algorithms combining hospitalizations, outpatient visits, and pharmaceutical records	Used Manitoba Primary Care Research Network (MaPCReN) EMR data as gold standard; developed multiple case definitions; evaluated sensitivity, specificity, AUROC, and Kappa across 1-, 2-, and 3-year windows
Ferrara et al. (2023)	Extract-Transform-Load (ETL) + Database Construction + Clustering Validation	Data engineering and unsupervised machine learning for mental health EHRs	Developed FEPSY database from Ferrara's mental health EHRs (46,222 patients, 3.8M records); anonymization via UUID replacement; clustering analysis using adapted k-means for categorical data to validate data structure; preliminary validation showed two distinct patient clusters
Zang et al. (2024)	Comparative Evaluation of Regularized Logistic Regression, GBM, and LSTM across Three Datasets (Claims + EHR)	Cross-dataset transportability analysis using real-world data	Models trained on APCD (claims), HIDD (inpatient EHR), and KHIN (all settings EHR); evaluated both local performance (same dataset) and transported performance (cross-dataset); predictors included demographics and diagnosis codes; hyperparameter tuning by cross-validation; significance tested via bootstrapping
Cillessen et al. (2017)	Retrospective System Usage Analysis + Cross-Sectional Survey (CSUQ)	System implementation evaluation combining log analysis and usability survey	Usage extracted from application event logs and EHR database; CSUQ adapted for satisfaction; stratified analysis by sex, age, specialty; statistical testing (ANOVA, Pearson correlation) to find factors

Confirming that administrative case definitions actually reflect patient reality is still key if researchers and public-health officials want to trust conclusions drawn from electronic health records. Schwartz et al. (2019) compared medication orders recorded in EMRs with real-world pharmacy-dispensing logs and found that the clinical systems consistently missed a sizable proportion of filled prescriptions. Nakamura et al. (2015) ran a parallel investigation of national influenza-reporting networks and cautioned that sentinel clinics tended to exaggerate weekly case counts. Ammann et al. (2018) reviewed clinical charts linked to ICD-9-CM codes for venous thromboembolism and showed that validation rates fluctuated sharply depending on where in the discharge record the diagnosis was entered. Dart et al. (2023) and team tested administrative definitions of paediatric hypertension and highlighted the

compromise between catching every true case and inadvertently labelling healthy children as patients, a trade-off shaped by how far back the data retrieval interval extends. Together, these investigations call for methodical algorithm building and rigorous record-validation steps whenever EHRs underpin large-scale epidemiology.

A few studies performed systematic literature reviews or developed best practice documents aiming to optimise future research and practices around EHR implementation. ASGE Technology Committee et al. (2016) mapped functionalities across endoscopic EMR vendors, providing a reference point for integration and quality verification. Wang & Zhang (2024) reviewed the healthcare uses of large language models (LLMs), pinpointing critical challenges and opportunities in AI clinical integration. Deignan et al. (2019) established

expert consensus on reanalysis procedures of genomic test data, thus ensuring frameworks for updating clinical relevance. Pacurariu et al. (2018) critically analysed the regulatory frameworks of several EHR-derived databases concerning their use for pharmacovigilance, underscoring cognitive trust requirements along with validation and accessibility benchmarks needed for secondary use. With increasing concerns regarding data privacy and security, blockchain technologies are being researched to guarantee secure EMR management. Regueiro et al. (2021) proposed an architecture that connects local EMR databases to Ethereum-based blockchains through a synchroniser mechanism, ensuring resilience during partial network failures. Their evaluation showed that blockchain-based synchronisation maintained real-time data integrity despite latency and throughput challenges, paving the way for secure interoperable systems in health information technology.

Challenges of EHR

Alongside defining the methodological frameworks, this review sequentially extracted and

categorized the principal challenges reported in all 32 selected studies concerning the design, implementation, evaluation, and application of electronic health record (EHR) systems. A comprehensive summary of these challenges is found in Table 2, which elucidates the persistent technical, organisational, and user-centred obstacles encountered. The studies reveal recurrent issues, including data privacy concerns as noted by Xanthidou et al. (2018) and Kehl et al. (2024). Interoperability and inconsistency in terminologies were highlighted by Zhang et al. (2022) and Pacurariu et al. (2018). System usability issues were examined by Jokar et al. (2023) and Burt et al. (2017). Data incompleteness and quality deficits were emphasized by Muñoyerro-Muñiz et al. (2020) and Huber et al. (2022). Regarding scalability, with respect to infrastructure and machine learning, limitations were reported by Cho et al. (2022), Xiang et al. (2019), and Cui et al. (2023). In addition, several studies emphasise resource limitations, concerns regarding generalizability, and regulatory frameworks that hinder the sustainability and wider applicability of EHR initiatives.

Table 2. Challenges

Authors (Year)	Identified Challenges	Brief Description
Xanthidou et al. (2018)	Data privacy and sensitivity, economic feasibility, interoperability, increased administrative burden on doctors	Privacy: categorization of sensitive demographic data; Economy: EMR costs vs. health system budgets; Technical: system interoperability; User: doctors' time spent on data input
Zhang et al. (2022)	Terminology inconsistency, difficulty in standardizing clinical knowledge, dependency on corpus quality, limited generalizability	Complex natural language descriptions required standardization; accuracy highly dependent on quality and size of the training corpus; single hospital data limits broad applicability
Jokar et al. (2023)	System usability issues, user adaptation time, incomplete initial data entry, limited scalability	Users initially faced difficulty in fully adapting to the new system; limited trial period may affect usability assessment; scalability beyond single center not tested
Xiang et al. (2019)	Temporal sparsity, standardization inconsistency, model dependency on time window size, limited generalizability to lab tests	Need for better handling of sparse temporal sequences; difficulty in normalizing diverse clinical codes; generalizability issues outside of diseases, diagnoses, and procedures due to data variation
Regueiro et al. (2021)	Network connectivity failures, blockchain synchronization delays, scalability limitations, high storage requirements	Disconnections impact blockchain update latency; high data growth rates challenge storage; proposed improvements include portable media backup and new data structures for more efficient queries
Muñoyerro-Muñiz et al. (2020)	Incomplete data integration, data quality inconsistency, ETL process complexity, metadata governance, real-time information delivery delays	Gaps in system coverage (e.g., outpatient data), difficulties ensuring data consistency across systems, complexity in updating ETL pipelines and metadata management, challenge of providing real-time analytics for healthcare management
Cho et al. (2022)	Data heterogeneity, outcome imbalance, graph storage complexity, limited disease generalization	Challenge in representing multi-type EMR entities; difficulty training with imbalanced positive/negative outcomes; storage management for large graph data; lack of validation across different disease groups outside CVD
Allorto & Wise (2015)	User training dependency, data omission due to non-mandatory fields, limited scalability, single-developer dependency	High reliance on user compliance for data completeness; no mandatory fields led to missing data; single developer risk for maintenance; plans needed for broader staff training and system expansion

ASGE Technology Committee et al. (2016)	Lack of EEMR standardization, complex EHR interoperability, high costs, dependency on IT infrastructure, variable user acceptance	Interoperability depends heavily on HL7 compliance and vendor cooperation; training needs; cost burden for small practices; absence of standardized data search tools across systems; limitations in free text querying
Burt et al. (2017)	Lack of standardization in usability assessment tools for EMR visualizations; heterogeneity in visualization formats; risk of poor data display usability impacting clinical decision-making	Variability in existing methods; fragmented usability metrics; challenge in ensuring universally high usability across different EMR visualization systems
Pacurariu et al. (2018)	Data fragmentation, lack of interoperability, incomplete data capture (especially inpatient drugs, biomarkers), lack of standardized validation, limited database access	Regulatory-level limitations due to diverse database structures; difficulties in integrating data for multinational studies; challenges in accessing and validating secondary-use data; efforts needed to harmonize databases through CDM and improve transparency
Bonomi & Jiang (2018)	Data sparsity, temporal asynchrony, high dimensionality, need for domain-specific pattern selection, assumption of independence in null model	Temporal misalignment between patients, difficulty matching multi-event patterns across heterogeneous sequences, scalability challenges for very large temporal datasets, risk if initial user-defined patterns are suboptimal
Siefridt et al. (2020)	Context ambiguity, stemming errors, synonym mismatches, limited terminological translation coverage, annotation subjectivity	Misinterpretation due to ambiguous terms (e.g., "radio"), stemming artifacts (e.g., "grosse" annotated as "grosse"), partial terminology translations impact French corpus, subjective evaluator bias
Schwartz et al. (2019)	Under-capture of antibiotics in EMR, discrepancy between prescribed and dispensed data, reliance on non-primary care prescribers, limitations in tracking real patient adherence	EMR systems captured only a subset of true antibiotic consumption; significant number of patients filled prescriptions from non-EMRALD physicians; differences between written and filled prescriptions complicate surveillance; generalizability limited to older adults
Nakamura et al. (2015)	Overestimation by sentinel surveillance, delay in reporting, non-random sentinel selection bias, pharmacy-based estimation limitations	NOSSID's overestimation due to voluntary sentinel selection; reporting delays of 7-10 days; PS reliance on pharmacy reporting coverage; incomplete matching of influenza season dates in datasets
Wang & Zhang (2024)	Data privacy, inaccurate output ("hallucinations"), fairness and bias, copyright concerns, limited multilingual adaptation, regulatory compliance gaps	LLMs face data leakage risks, often generate plausible but incorrect information, show demographic biases, face plagiarism issues, and struggle with adapting to non-English healthcare systems
Deignan et al. (2019)	Resource constraints for periodic reevaluation, lack of reimbursement models, risk of outdated variant classifications, responsibility ambiguities between labs and clinicians	Reevaluation requires considerable manpower and costs without guaranteed payment; responsibility for initiating reevaluation may be unclear; necessity of robust internal variant databases and version tracking; frequent updates needed for consistency with new ACMG standards
Ammann et al. (2018)	Incomplete chart retrieval, low PPV for position-unspecified codes, time lag between symptom onset and recorded admission, patient population specificity (IGIV users)	Difficulty in matching administrative data with real clinical events; lower reliability of physician billing claims vs hospital claims; gap in accuracy when onset occurred before hospitalization; findings primarily reflect IGIV patient cohort
Huber et al. (2022)	Inconsistent register use, incomplete ID recording, gaps in linkage of lab data, duplication of documentation efforts, variable non-standard registers	Register rationalization incomplete; high administrative burden on clinic staff; difficulty tracking patients longitudinally due to missing IDs; dependence on paper forms in many non-HIV conditions; risk of data flow breakdown
Sander & Wauer (2019)	Dependency on terminology server quality, limitations in handling nested queries, historical data semantic gaps, scalability for large-	System performance depends heavily on coverage and accuracy of the terminology server; difficulties arise in historical datasets with outdated or incomplete medical terms; scaling O-SQL to extremely large databases could

Parejo et al. (2020)	scale queries Scalability for large tag clouds, limited to short text records, dependence on preprocessing quality (synonym, acronym resolution), display resolution limits	introduce latency unless optimization strategies are applied Difficulty representing large heterogeneous datasets visually; limited testing on long clinical narratives; system performance highly dependent on preprocessing accuracy (synonym/acronym replacement); visual overload risk on small screens
Cui et al. (2023)	Dependency on quality and availability of similar patients, sensitivity to hyperparameters (number of similar patients and look-back steps), scalability issues for extremely large datasets	Performance varies with patient similarity quality; requires careful tuning of number of similar patients (K=10 optimal); difficulty in federated settings where patient data sharing is restricted; challenges in extending method to other diseases without retraining
Freire et al. (2016)	Slow response times for XML databases, large storage overhead for Couchbase, high indexing costs, slower scaling performance with more Couchbase nodes	XML-based NoSQL systems not scalable for population-based queries; Couchbase indexing is query-specific and expensive for large datasets; cluster expansion (more nodes) in Couchbase adds coordination delays if network is slow; generalization limited to archetype-modeled EHR
Kim et al. (2021)	Low standard vocabulary coverage, complexity of PSG data structures, manual effort for concept mapping, site variability in PSG data capture, resource-intensive CDM conversion	Only 23% of parameters matched standard vocabularies, requiring custom concepts; EHR variability across sites challenges harmonization; high initial investment in time and resources for full PSG data CDM conversion; dependency on collaboration across clinical, terminology, and IT experts
Nash et al. (2017)	Data quality and completeness issues in EMR, limited generalizability beyond PS Suite users, possible contamination between clusters, limited adverse event capture	Use of real-world EMR data introduces noise; only indirect adverse event capture; generalizability restricted to voluntary EMERALD users; need for high physician engagement with tools; tool adoption variability
Nash et al. (2017)	Variability in EMR data quality, incomplete capture of adverse events, dependence on voluntary physician engagement, generalizability issues beyond PS Suite users	EMR records may miss key events (e.g., statin side effects); intervention effectiveness dependent on actual physician adoption; findings may not generalize to EMR systems beyond PS Suite; contamination possible as control clinics may adopt similar practices
Alzoubi et al. (2019)	EHR data noise and missingness, lack of standardization in phenotyping definitions, complexity of unstructured data extraction, difficulty creating gold standards, variable access to EHRs	Clinical data heterogeneity and missing data impede feature extraction; manual labelling for gold standards resource-intensive; inconsistent availability of narrative EHR data; phenotyping algorithms often site-specific and non-generalizable; access restrictions to EHR data
Kehl et al. (2024)	Risk of PHI leakage in teacher models, dependency on public dataset quality for student training, generalization limitations beyond cancer domain, dataset bias toward academic centers	Direct training on PHI introduces membership inference risk; performance of student models sensitive to quality and representativeness of public datasets (MIMIC-IV); teacher-student approach preserves privacy but external validity to smaller hospitals remains untested; limited disease coverage
Dart et al. (2023)	Low sensitivity of administrative algorithms, under-recording of BP in EMRs, selection bias from children with BP recorded, limited generalizability to other regions	Administrative algorithms underdetect true cases without pharmaceutical data; only 26.7% of clinic visits had BP recorded; different healthcare recording practices in other regions may affect replicability; updating gold standard needed as guidelines evolve
Ferrara et al. (2023)	Data heterogeneity from multiple EHR systems, missing data, privacy and anonymization challenges, generalizability issues due to localized sample	Integration of legacy data from 6 different EHR systems required extensive cleaning; anonymization removed potentially useful fields (e.g., full date of birth); missing sociodemographic information in many records; findings based on a single health district (Ferrara province)
Zang et al. (2024)	Performance degradation when models transported across datasets, data heterogeneity, feature	Clinical context differences between claims and EHR datasets impacted model generalizability; LSTM overfit in source datasets without improving transported

Cillessen et al. (2017)	prevalence differences, model overfitting, limited feature standardization Variations in user satisfaction across specialties, insufficient awareness of training, limited coding usage (ICD-10), documentation burden concerns	performance; key predictors varied in importance across datasets; suggests need for adaptive or ensemble transport strategies Specialty-specific satisfaction differences suggest need for customization; half of physicians unaware of training opportunities; ICD-10 coding rarely used despite being available; documentation burden vs usability remains a tension; log file proved crucial for tracking system usage and troubleshooting
-------------------------	--	--

As seen in Table 2, some challenges of HER implementation were discovered. Those challenges are reported as follows. The issue of incomplete and inconsistent clinical data persists in the evaluation and implementation of EHR systems. In studies conducted by Muñoyerro-Muñiz et al. (2020) and Huber et al. (2022), the problems caused by data gaps (either through incomplete linkage or inconsistent recording practices) are shown to compromise the integrity of health data warehouses and patient registries. Dart et al. (2023) and Ferrara et al. (2023) have documented similar concerns whereby localised samples, undocumented covariates, and data loss spanning several EHR systems severely weaken the generalisability of the findings. Alzoubi et al. (2019) and Schwartz et al. (2019) highlight the noise and incompleteness within the EHR datasets, hampering attempts to set gold standards for case definitions and phenotyping. Furthermore, as shown by Nash et al. (2017) and Ammann et al. (2018), administrative datasets systematically misrecord important clinical phenomena, claiming to capture real clinical events while flagging symptoms and disguising them, which creates unresolvable biases. All these studies illuminate the urgent requirement for enhanced data stewardship, underscoring the need for assurance of completeness alongside stringent verification processes concerning data accuracy. These issues become increasingly important when EHR data are used for clinical research and clinical decision-making.

Like many other EHR systems, these systems are affected by the lack of universal interoperability standards as well as the fragmentation of clinical terminologies. Zhang et al. (2022) point out the issues with integrating unstructured clinical narratives due to monolingual and multilingual knowledge representation standards, which Siefridt et al. (2020) also express concern about regarding terminology inconsistency across languages. Other authors, such as the ASGE Technology Committee et al. (2016) and Pacurariu et al. (2018), have noted that intertwined gaps with systemic interoperability manifest not only in cross-institutional data exchange but also in the regulatory sufficiency of datasets derived from EHRs for secondary use. Furthermore, Sander & Wauer (2019) explain how SQL-based querying systems become

semantically limited in their scalability due to a reliance on external terminology servers. Similar obstacles were faced by Kim et al. (2021) in their effort to standardise CDM-converted complex polysomnographic data, resulting in the need to devise custom concepts. These persistent challenges signal an urgent call for enhanced terminology maps, unified standards, stronger interoperability frameworks, as well as the need and the primary aim of enabling diverse integration and secondary utilisation of health data.

Besides the technical specifications, the usability of EHR systems has a profound impact on their acceptance and effectiveness in each clinical setting. As also illustrated by Jokar et al. (2023) and Allorto & Wise (2015), even well-designed systems suffer from user resistance due to steep learning curves and workflow gridlocks. Cillessen et al. (2017) also noted that user satisfaction with EHRs during implementation had multidisciplinary gaps within specialties, reflecting under-training and poor design focus tailoring. Attempts to improve usability, as explored by Parejo et al. (2020) with graphical exploration tools and Burt et al. (2017) with visualisation evaluation frameworks, face stubborn barriers of diverse user preferences and clinical practices. Nash et al. (2017) point out that successful adoption is intricately linked not just to organisational systems but to their organisational support structures, describing tool deployment by stating that “Passive engagement strategies are not enough.” Collectively, these findings reaffirm the central importance of human factors engineering and embedded usability in EHR systems design and strategies for development and implementation.

Scalability, System Performance, and Technical Bottlenecks

As the scope and complexity of EHR systems increase, their scalability and technical issues take center stage. Regueiro et al. (2021) illustrate how synchronisation mechanisms using blockchain technology, despite having solid security claims, struggle with latency and storage scalability in practical networking contexts. Freire et al. (2016) show that XML-based traditional databases do not perform well with population-level queries. Xiang et al. (2019) and Cho et al. (2022) show that the management of extremely large,

diverse clinical datasets, whether via temporal embedding or graph-based techniques, poses storage and computational burdens far beyond the already limited infrastructure. Sander & Wauer (2019) go further to note that querying semantically at scale is dependent on the performance of the terminology server, making it quite fragile. Parejo et al. (2020) and Cui et al. (2023) also highlight the impact that scalability exerts on data visualisation, machine learning model training, and the latter's deployment processes. These studies emphasise the overriding need to pair architectural design and resource optimisation with performance-sensitive systems alongside technical advancements to prevent degradation as systems scale.

Data Privacy, Security Risks, and Ethical Concerns

The threats of a breach of data ethics, security, and the erasure of sensitive information have grown in tandem with the rise of digitized health records. Xanthidou et al. (2018) and Wang & Zhang (2024) explore the balance between obtaining comprehensive clinical information while also safeguarding patient confidentiality, particularly with the high-tech methods available today for reidentification. Kehl et al. (2024) showcase how even privacy-preserving AI techniques like teacher-student model distillation face challenges from membership inference attacks and biases from public datasets. Ferrara et al. (2023) show how attempts to comply with regulations on anonymisation, although necessary, can dangerously remove critical sociodemographic variables from datasets, hence tarnishing their usefulness for research. All these examples highlight the need for more dynamic and advanced privacy regulations that facilitate the use of data while encrypting sensitive information with advanced cryptographic methods, federated learning frameworks, and clear governance policies.

The employment of machine learning on EHR data has revealed both extraordinary promise and deeply concerning limitations, especially about model generalizability. As noted by Xiang et al. (2019) and Cho et al. (2022), model performance diminishes outside the data domain where they were trained, reinforcing the classic overfitting and temporal transferability pitfalls. Cui et al. (2023) and Zang et al. (2024) examine dataset and institution cross portability, showing that even the most advanced models suffer from degradation with shifting clinical contexts. Kehl et al. (2024) also describe AI systems trained on narrowly defined oncology datasets as likely to struggle when applied to broader, less curated, or multidisciplinary datasets. Taken together, these insights suggest that, at this point in time at least, external validity remains the last unfinished feature in clinical machine learning, demanding additional validation across population, disease, and

institutional diversity before bypassing careful algorithmic calibration.

Ultimately, even the most sophisticated technological solutions fail when the issues of organization, economy, and politics are left unattended. Research conducted by Xanthidou et al. (2018) and Deignan et al. (2019) demonstrates that also within the EHR systems, staffing, funding, and reimbursement structures are particularly daunting resource constraints that make sustaining EHR innovation difficult. Pacurariu et al. (2018) as well as the ASGE Technology Committee et al. (2016) depict how disjointed regulatory environments as well as gaps in infrastructure increase the difficulty of interoperability and adoption. At the same time, Nakamura et al. (2015) and Wang & Zhang (2024) illustrate the inconsistency of public health surveillance combined with regulatory inaction, which complicates the standardisation of data. This research reinforces the notion that the swift pace of technological improvements must be paralleled with comprehensive policy strategies, dependable funding, and institutional frameworks that can support EHR system changes.

Discussion

A survey of evaluation methods across the reviewed articles illustrates both the technical maturity and the continuing practical hurdles that shape research on electronic health records (EHRs). Studies focused on usability and implementation, including Jokar et al. (2023), Allorto & Wise (2015), Nash et al. (2017), and Cillessen et al. (2017), found that systems built to specification often fail to blend smoothly with daily clinical routines, leaving adoption rates uneven and revealing a persistent gap between intended and actual use. Meanwhile, the tricky task of stitching together disparate data streams for research-ready warehouses was tackled by Muñozerro-Muñiz et al. (2020), Freire et al. (2016), Kim et al. (2021), and Ferrara et al. (2023), who catalogued the ETL bottlenecks that still slow progress. Regueiro et al. (2021) then introduced blockchain as a potential synchronisation layer, yet that solution remains in the experimental stage and has not yet passed the test of broad real-world deployment (Negro-Calduch et al., 2021; Xiao et al., 2018).

Zhang et al. (2022), Siefridt et al. (2020), Bonomi & Jiang (2018), and Kehl et al. (2024) each highlighted recent gains in natural language processing, converting free-text clinical notes into usable, structured data. Despite the progress, researchers still worry about inconsistent data quality, lingering annotation biases, and the difficulty of fine-tuning such models for narrow clinical specialties (Cahill et al., 2025; Supriyadi et al., 2023; Dewanto et al., 2024; Winiasri et al., 2023). A broader survey on machine-learning applications, aimed at prediction, phenotyping, and system

modelling, appears in the work of Xiang et al. (2019), Cho et al. (2022), Cui et al. (2023), Alzoubi et al. (2019), and Zang et al. (2024). Their findings are mixed: although well-built predictive tools can strengthen clinical decision support, most fail to generalise easily across facilities, and many simpler models outperform elaborate networks when tested outside the training environment (Torab-Miandoab et al., 2023). Schwartz et al. (2019), Nakamura et al. (2015), Ammann et al. (2018), and Dart et al. (2023) each examined the reliability of administrative databases and cautioned that persistent underreporting and inconsistent coding still limit their value in surveillance and epidemiology. In parallel, the ASGE Technology Committee et al. (2016), along with Wang & Zhang (2024), Deignan et al. (2019), and Pacurariu et al. (2018), conducted systematic reviews that packaged the evidence into practical guidelines and thematic policy recommendations. Regueiro et al. (2021) then explored early-stage blockchain prototypes for electronic health record management, framing them as promising yet unfinished infrastructure solutions. Taken together, this body of work portrays a rapidly evolving landscape where exciting technological advances are accompanied by lingering gaps in standardisation, validation, user acceptance, and ethical oversight (Koleck et al., 2019; Reegu et al., 2023)

A review of the studies collected here reveals that many distinct problems still block effective adoption and everyday use of electronic health record (EHR) systems. Data quality, completeness, and missing information continue to trouble researchers, as Muñozerro-Muñiz et al. (2020), Huber et al. (2022), and Dart et al. (2023) show; inconsistent entry habits, unrecorded fields, and broken data routes together weaken the credibility and wider usefulness of evidence drawn from EHRs (Abramson et al., 2012). The absence of unified terms and the stubborn gap between different software impede progress, with Zhang et al. (2022), Pacurariu et al. (2018), and Sander & Wauer (2019) illustrating that missing standard ontologies and common exchange rules slow integration between systems and reuse of pooled information (Meystre et al., 2008). Finally, usability issues raised by Jokar et al. (2023), Allorto & Wise (2015), and Cillessen et al. (2017) remind us that even the most powerful technology will fail in busy clinics if staff are not trained properly, screens are poorly designed, and system tasks do not match routine work. Several recent studies – Regueiro et al. (2021), Freire et al. (2016), Cho et al. (2022), and Xiang et al. (2019) – independently document similar scalability and performance pitfalls, showing that system infrastructure buckles under heavy loads, queries return slowly, and state-of-the-art large-scale models demand so much compute that running them routinely jeopardises long-term sustainability (Cook et

al., 2022). At the same time, Kehl et al. (2024) and Wang & Zhang (2024) raise urgent ethical questions, arguing that every effort to make patient data more available for AI-driven analysis risks breaching confidentiality, a tension that federated learning was supposed to ease but does not fully resolve. Generalisability problems receive fresh attention from Zang et al. (2024) and Cui et al. (2023), who observe that models trained in one hospital often fail or degrade when tested in another clinical setting, pointing to the pressing need for external validation protocols that go beyond the familiar holdout dataset (Goldstein et al., 2017; Lewis et al., 2023; Reegu et al., 2023; Putra et al., 2023; Santosa et al., 2020; Zulkifli et al., 2022)

Importantly, the technical and methodological issues discussed cannot be understood in isolation from the wider systemic and organisational problems that shape health information platforms. Xanthidou et al. (2018) have each noted enduring roadblocks such as scarce institutional funding, shifting regulatory criteria, and fragmented governance, all of which together undermine the sustainability and spread of electronic health record systems. These systemic shortfalls also echo the limits of the evidence surveyed here, since many original studies drew data from a single hospital, tested prototypes only in lab settings, or failed to track outcomes over extended periods, curtailing how widely their conclusions can be applied. In addition, the review process itself may have introduced selection bias through differential inclusion of published work, and the range of designs reviewed adds heterogeneity that complicates the synthesis and interpretation of results (V.H. et al., 2010).

The evidence reviewed suggests that any ambition to advance electronic health records from technical novelty to reliable health system workhorse hinges on four interdependent pillars: resilient infrastructure, a skilled and continually trained workforce, updated legal and governance frameworks, and design that places patients at the centre. Translating this insight into day-to-day practice means habitually integrating usability testing and clinical validation at the outset of each development sprint. At the policy level, regulators and funders must synchronise resources behind open technical standards and commitments to affordable, equitable access. On the research front, investigators should prioritise multicentre, longitudinal studies that monitor performance in diverse settings and advance ethical artificial intelligence anchored to clinicians' real workloads. Collectively, these steps advance Sustainable Development Goals analogous to good health (SDG 3), to resilient infrastructure (SDG 9), and to accountable institutions (SDG 16) and thus map a robust, equity-driven digital health strategy (Haque et al., 2022; Aliabadi et al., 2020).

Conclusion

The systematic review shows that researchers have tested electronic health record (EHR) systems using a surprisingly broad set of methods, underscoring the field's mix of qualitative and quantitative approaches. Some teams employed phenomenological interviews and standard usability testing; others built machine-learning models, benchmarked database queries, examined blockchain architectures, or compiled evidence in formal literature reviews, each technique chosen to probe specific questions about performance, user experience, integration, or predictive power. Despite this diversity, a cluster of long-standing obstacles appeared again and again. Data quality, completeness, and missingness still undermine the trustworthiness of retrospective analyses; clashing vocabularies and weak interoperability slow data exchange between hospitals and frustrate regulators; confusing interfaces and steep learning curves keep many clinicians from fully engaging; ageing hardware and software architecture choke scalability; heightened analytics expose new privacy and security fears; and policy uncertainties, budget limits, and shifting priorities tug at the sustainability of any EHR initiative. Taken together, these insights indicate that the methodological toolbox for EHR studies has advanced, yet real progress now depends on tackling these interlocking systemic barriers so that promising technologies deliver concrete gains in patient care and public health.

Acknowledgments

The author would like to thank the parties who have contributed to this research.

Author Contributions

In this study, there were two researchers, namely Ade Dita Puteri who contributed to collecting data, analyzing and interpreting and Lira Mufti Azzahri Isnaeni who contributed to filtering and selecting research data.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Abramson, E. L., McGinnis, S., Edwards, A., Maniccia, D. M., Moore, J., & Kaushal, R. (2012). Electronic Health Record Adoption and Health Information Exchange Among Hospitals in New York State. *Journal of Evaluation in Clinical Practice*, 18(6), 1156–1162. <https://doi.org/10.1111/j.1365-2753.2011.01755.x>
- Aliabadi, A., Sheikhtaheri, A., & Ansari, H. (2020). Electronic Health Record-Based Disease Surveillance Systems: A Systematic Literature Review on Challenges and Solutions. *Journal of the American Medical Informatics Association*, 27(12), 1977–1986. <https://doi.org/10.1093/jamia/ocaa186>
- Allorto, N. L., & Wise, R. D. (2015). Development and Evaluation of an Integrated Electronic Data Management System in a South African Metropolitan Critical Care Service. *Southern African Journal of Anaesthesia and Analgesia*, 21(2), 31–35. <https://doi.org/10.1080/22201181.2015.1115607>
- Alqahtani, A., Crowder, R., & Wills, G. (2017). Barriers to the Adoption of EHR Systems in the Kingdom of Saudi Arabia: An Exploratory Study Using a Systematic Literature Review. *Journal of Health Informatics in Developing Countries*, 11(2). Retrieved from <https://www.jhidc.org/index.php/jhidc/article/view/160>
- Alzoubi, H., Alzubi, R., Ramzan, N., West, D., Al-Hadhrani, T., & Alazab, M. (2019). A Review of Automatic Phenotyping Approaches Using Electronic Health Records. *Electronics*, 8(11), 1235. <https://doi.org/10.3390/electronics8111235>
- Ammann, E. M., Cuker, A., Carnahan, R. M., Perepu, U. S., Winiecki, S. K., Schweizer, M. L., ... Beukelman, T. (2018). Chart Validation of Inpatient ICD-9-CM Administrative Diagnosis Codes for Venous Thromboembolism Among Intravenous Immune Globulin Users. *Medicine*, 97(38), e9960. <https://doi.org/10.1097/MD.00000000000009960>
- ASGE Technology Committee, Manfredi, M. A., Chauhan, S. S., Enestvedt, B. K., Fujii-Lau, L. L., Konda, V., Hwang, J. H., Lo, S. K., Maple, J. T., Murad, F. M., Pannala, R., Woods, K. L., & Banerjee, S. (2016). Endoscopic Electronic Medical Record Systems. *Gastrointestinal Endoscopy*, 83(1), 29–36. <https://doi.org/10.1016/j.gie.2015.09.036>
- Bonomi, L., & Jiang, X. (2018). Patient Ranking with Temporally Annotated Data. *Journal of Biomedical Informatics*, 78, 43–53. <https://doi.org/10.1016/j.jbi.2017.12.007>
- Burt, A., Morgan, L., Petrinic, T., Young, D., & Watkinson, P. (2017). Usability Evaluation Methods Employed to Assess Information Visualisations of Electronically Stored Patient Data for Clinical Use: A Protocol for a Systematic Review. *Systematic Reviews*, 6, Article 111. <https://doi.org/10.1186/s13643-017-0544-1>
- Cahill, M., Cleary, B. J., & Cullinan, S. (2025). The Influence of Electronic Health Record Design on Usability and Medication Safety: Systematic

- Review. *BMC Health Services Research*, 25(1). <https://doi.org/10.1186/s12913-024-12060-2>
- Cho, H. N., Ahn, I., Gwon, H., Kang, H. J., Kim, Y., Seo, H., ... Kim, S. (2022). Heterogeneous Graph Construction and HinSAGE Learning from Electronic Medical Records. *Scientific Reports*, 12, Article 20933. <https://doi.org/10.1038/s41598-022-25693-2>
- Cillessen, F. H. J. M., Robbé, P. F. D. V., & Biermans, M. C. J. (2017). A Hospital-Wide Transition from Paper to Digital Problem-Oriented Clinical Notes: A Descriptive History and Cross-Sectional Survey of Use, Usability, and Satisfaction. *Applied Clinical Informatics*, 8(2), 502-514. <https://doi.org/10.4338/ACI-2016-08-RA-0137>
- Cook, L. A., Sachs, J., & Weiskopf, N. G. (2022). The Quality of Social Determinants Data in the Electronic Health Record: A Systematic Review. *Journal of the American Medical Informatics Association*, 29(1), 187-196. <https://doi.org/10.1093/jamia/ocab199>
- Cui, S., Wei, G., Zhou, L., Zhao, E., Wang, T., & Ma, F. (2023). Predicting Line of Therapy Transition via Similar Patient Augmentation. *Journal of Biomedical Informatics*, 147, 104511. <https://doi.org/10.1016/j.jbi.2023.104511>
- Dart, A., Singer, A., Chanchlani, R., Ferguson, T., Tangri, N., & Zappitelli, M. (2023). Evaluation of Administrative Case Definitions for Hypertension in Canadian Children. *Scientific Reports*, 13, Article 33401. <https://doi.org/10.1038/s41598-023-33401-x>
- Deignan, J. L., Chung, W. K., Kearney, H. M., Monaghan, K. G., Rehder, C. W., & Chao, E. C. (2019). Points to Consider in the Reevaluation and Reanalysis of Genomic Test Results: A Statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, 21(6), 1267-1270. <https://doi.org/10.1038/s41436>
- Dewanto, D., Santosa, T. A., Ratih, A., Asrizal, A., & Hardeli, H. (2024). The Influence of the STEM-Based Guided Inquiry Model on Students' Creative Thinking Skills in Science Learning: A Meta-Analysis Study. *Jurnal Penelitian Pendidikan IPA*, 10(3), 88-95. <https://doi.org/10.29303/jppipa.v10i3.6777>
- Elfira, I., & Santosa, T. A. (2023). Literature Study: Utilization of the PjBL Model in Science Education to Improve Creativity and Critical Thinking Skills. *Jurnal Penelitian Pendidikan IPA*, 9(1), 133-143. <https://doi.org/10.29303/jppipa.v9i1.2555>
- Ettaloui, N., Arezki, S., & Gadi, T. (2024). Blockchain-Based Electronic Health Record: Systematic Literature Review. *Human Behavior and Emerging Technologies*, 2024(1). <https://doi.org/10.1155/hbe2/4734288>
- Ferrara, M., Gentili, E., Murri, M. B., Zese, R., Alberti, M., Franchini, G., ... Amore, M. (2023). Establishment of a Public Mental Health Database for Research Purposes in the Ferrara Province: Development and Preliminary Evaluation Study. *JMIR Medical Informatics*, 11, e45523. <https://doi.org/10.2196/45523>
- Freire, S. M., Teodoro, D., Wei-Kleiner, F., Sundvall, E., Karlsson, D., & Lambrix, P. (2016). Comparing the Performance of NoSQL Approaches for Managing Archetype-Based Electronic Health Record Data. *PLOS ONE*, 11(3), e0150069. <https://doi.org/10.1371/journal.pone.0150069>
- Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. A. (2017). Opportunities and Challenges in Developing Risk Prediction Models with Electronic Health Records Data: A Systematic Review. *Journal of the American Medical Informatics Association*, 24(1), 198-208. <https://doi.org/10.1093/jamia/ocw042>
- Haque, A. K. M. B., Arifuzzaman, B. M., Siddik, S. A. N., Kalam, A., Shahjahan, T. S., Saleena, T. S., Alam, M., Islam, M. R., Ahmmed, F., & Hossain, M. J. (2022). Semantic Web in Healthcare: A Systematic Literature Review of Application, Research Gap, and Future Research Avenues. *International Journal of Clinical Practice*, 2022. <https://doi.org/10.1155/2022/6807484>
- Huber, A. N., Pascoe, S., Fox, M. P., Murphy, J., Phokojoe, M., Gorgens, M., ... Fairall, L. (2022). Leveraging Routine Data in Impact Evaluation: Understanding Data Systems in Primary Healthcare Prior to a Matched Cluster-Randomised Evaluation of Adherence Guidelines in South Africa. *South African Medical Journal*, 112(10), 819-827. <https://doi.org/10.7196/SAMJ.2022.v112i10.14909>
- Ichsan, I., Suharyat, Y., Santosa, T. A., & Satria, E. (2023). Effectiveness of STEM-Based Learning in Teaching 21st Century Skills in Generation Z Student in Science Learning: A Meta-Analysis. *Jurnal Penelitian Pendidikan IPA*, 9(1), 150-166. <https://doi.org/10.29303/jppipa.v9i1.2517>
- Jokar, M., Sahmeddini, M. A., Zand, F., Rezaee, R., & Bashiri, A. (2023). Development and Evaluation of an Anesthesia Module for Electronic Medical Records in the Operating Room: An Applied Developmental Study. *BMC Anesthesiology*, 23, Article 335. <https://doi.org/10.1186/s12871-023-02335-2>
- Katurura, M. C., & Cilliers, L. (2018). Electronic Health Record System in the Public Health Care Sector of South Africa: A Systematic Literature Review.

- African Journal of Primary Health Care & Family Medicine*, 1-8. <https://doi.org/10.4102/>
- Kehl, K. L., Jee, J., Pichotta, K., Paul, M. A., Trukhanov, P., Fong, C., ... Weissman, J. S. (2024). Shareable Artificial Intelligence to Extract Cancer Outcomes from Electronic Health Records for Precision Oncology Research. *Nature Communications*, 15, Article 9787. <https://doi.org/10.1038/s41467-024-54071-x>
- Kim, J. W., Kim, S., Ryu, B., Song, W., Lee, H. Y., & Yoo, S. (2021). Transforming Electronic Health Record Polysomnographic Data into the OMOP Common Data Model: A Pilot Feasibility Study. *Scientific Reports*, 11, Article 86564. <https://doi.org/10.1038/s41598-021-86564-w>
- Koleck, T. A., Dreisbach, C., Bourne, P. E., & Bakken, S. (2019). Natural Language Processing of Symptoms Documented in Free-Text Narratives of Electronic Health Records: A Systematic Review. *Journal of the American Medical Informatics Association*, 26(4), 364-379. <https://doi.org/10.1093/jamia/ocy173>
- Kruse, C. S., Kristof, C., Jones, B., Mitchell, E., & Martinez, A. (2016). Barriers to Electronic Health Record Adoption: a Systematic Literature Review. *Journal of Medical Systems*, 40(12). <https://doi.org/10.1007/s10916-016-0628-9>
- Lewis, A. E., Weiskopf, N., Abrams, Z. B., Foraker, R., Lai, A. M., Payne, P. R. O., & Gupta, A. (2023). Electronic Health Record Data Quality Assessment and Tools: A Systematic Review. *Journal of the American Medical Informatics Association*, 30(10), 1730-1740. <https://doi.org/10.1093/jamia/ocad120>
- Mayer, A. H., Costa, C. A. D., & Righi, R. D. R. (2020). Electronic Health Records in a Blockchain: A Systematic Review. *Health Informatics Journal*, 26(2), 1273-1288. <https://doi.org/10.1177/1460458219866350>
- Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F. (2008). Extracting Information from Textual Documents in the Electronic Health Record: A Review of Recent Research. *Yearbook of Medical Informatics*, 128-144. <https://doi.org/10.1055/s-0038-1638592>
- Muñoyerro-Muñiz, D., Goicoechea-Salazar, J. A., García-León, F. J., Laguna-Téllez, A., Larrocha-Mata, D., & Cardero-Rivas, M. (2020). Health Record Linkage: Andalusian Health Population Database. *Gaceta Sanitaria*, 34(2), 105-113. <https://doi.org/10.1016/j.gaceta.2019.03.003>
- Nakamura, Y., Sugawara, T., Kawano-hara, H., Ohkusa, Y., Kamei, M., & Oishi, K. (2015). Evaluation of Estimated Number of Influenza Patients from National Sentinel Surveillance Using the National Database of Electronic Medical Claims. *Japanese Journal of Infectious Diseases*, 68(1), 27-29. <https://doi.org/10.7883/yoken.JJID.2014.092>
- Nash, D. M., Ivers, N. M., Young, J., Jaakkimainen, R. L., Garg, A. X., & Tu, K. (2017). Improving Care for Patients with or at Risk for Chronic Kidney Disease Using Electronic Medical Record Interventions: A Pragmatic Cluster-Randomized Trial Protocol. *Canadian Journal of Kidney Health and Disease*, 4, 2054358117699833. <https://doi.org/10.1177/2054358117699833>
- Negro-Calduch, E., Azzopardi-Muscat, N., Krishnamurthy, R. S., & Novillo-Ortiz, D. (2021). Technological Progress in Electronic Health Record System Optimization: Systematic Review of Systematic Literature Reviews. *International Journal of Medical Informatics*, 152, 104507. <https://doi.org/10.1016/j.ijmedinf.2021.104507>
- Pacurariu, A., Plueschke, K., McGettigan, P., Morales, D. R., Slattery, J., Vogl, D., Goedecke, T., Kurz, X., & Cave, A. (2018). Electronic Healthcare Databases in Europe: Descriptive Analysis of Characteristics and Potential for Use in Medicines Regulation. *BMJ Open*, 8(9), e023090. <https://doi.org/10.1136/bmjopen-2018-023090>
- Parejo, U. T., Campaña, J. R., Vila, M. A., & Calvo-Flores, M. D. (2020). A Survey of Tag Clouds as Tools for Information Retrieval and Content Representation. *Information Visualization*, 20(10), 147387162096663. <https://doi.org/10.1177/1473871620966638>
- Payrovnaziri, S. N., Chen, Z., Rengifo-Moreno, P., Miller, T., Bian, J., Chen, J. H., Liu, X., & He, Z. (2020). Explainable Artificial Intelligence Models Using Real-World Electronic Health Record Data: A Systematic Scoping Review. *Journal of the American Medical Informatics Association*, 27(7), 1173-1185. <https://doi.org/10.1093/jamia/ocaa053>
- Putra, M., Rahman, A., Ilwandri, I., Suhayat, Y., Santosa, T. A., Putra, R., & Aprilisia, S. (2023). The Effect of STEM-Based REACT Model on Students' Critical Thinking Skills: A Meta-Analysis Study. *LITERACY: International Scientific Journals of Social, Education, Humanities*, 2(1), 207-217. <https://doi.org/10.56910/literacy.v2i1.560>
- Reegu, F. A., Abas, H., Gulzar, Y., Xin, Q., Alwan, A. A., Jabbari, A., Sonkamble, R. G., & Dziyauddin, R. A. (2023). Blockchain-Based Framework for Interoperable Electronic Health Records for an Improved Healthcare System. *Sustainability (Switzerland)*, 15(8). <https://doi.org/10.3390/su15086337>
- Regueiro, C., Seco, I., Gutiérrez-Agüero, I., Urquizu, B., & Mansel, J. (2021). A Blockchain-Based Audit Trail Mechanism: Design and Implementation. *Algorithms*, 14(12), 341.

- <https://doi.org/10.3390/a14120341>
- Sander, A., & Wauer, R. (2019). Integrating Terminologies into Standard SQL: A New Approach for Research on Routine Data. *Journal of Biomedical Semantics*, 10, Article 9. <https://doi.org/10.1186/s13326-019-0199-z>
- Santosa, T. A., Agustina, N., & Yulianti, S. (2020). Jenis Tumbuhan Liar dalam Upacara Adat Kenduri Seko di Kerinci. *Pendekar: Jurnal Pendidikan Berkarakter*, 3(1), 6–10. Retrieved from <http://journal.ummat.ac.id/index.php/pendekar/article/view/2798>
- Schwartz, K. L., Wilton, A. S., Langford, B. J., Brown, K. A., Daneman, N., Garber, G., ... Morris, A. M. (2019). Comparing Prescribing and Dispensing Databases to Study Antibiotic Use: A Validation Study of the Electronic Medical Record Administrative Data Linked Database (EMRALD). *Journal of Antimicrobial Chemotherapy*, 74(7), 2091–2097. <https://doi.org/10.1093/jac/dkz033>
- Sezgin, E., & Özkan, S. (2013). A Systematic Literature Review on Health Recommender Systems. *2013 E-Health and Bioengineering Conference, EHB 2013*. <https://doi.org/10.1109/EHB.2013.6707249>
- Shaikh, M., Vayani, A. H. M., Akram, S., & Qamar, N. (2022). Open-Source Electronic Health Record Systems: A Systematic Review of Most Recent Advances. *Health Informatics Journal*, 28(2). <https://doi.org/10.1177/14604582221099828>
- Shankar, R., Bunde, A., & Mukhopadhyay, A. (2025). Natural Language Processing of Electronic Health Records for Early Detection of Cognitive Decline: A Systematic Review. *NPJ Digital Medicine*, 8(1). <https://doi.org/10.1038/s41746-025-01527-z>
- Siefridt, C., Grosjean, J., Lefebvre, T., Rollin, L., Darmoni, S., & Schuers, M. (2020). Evaluation of Automatic Annotation by a Multi-Terminological Concepts Extractor within a Corpus of Data from Family Medicine Consultations. *International Journal of Medical Informatics*, 133, 104009. <https://doi.org/10.1016/j.ijmedinf.2019.104009>
- Supriyadi, A., Desy, D., Suharyat, Y., Santosa, T. A., & Sofianora, A. (2023). The Effectiveness of STEM-Integrated Blended Learning on Indonesia Student Scientific Literacy: A Meta-Analysis. *International Journal of Education and Literature*, 2(1), 41–48. <https://doi.org/10.55606/ijel.v2i1.53>
- Tapuria, A., Porat, T., Kalra, D., Dsouza, G., Xiaohui, S., & Curcin, V. (2021). Impact of Patient Access to Their Electronic Health Record: Systematic Review. *Informatics for Health and Social Care*, 46(2), 192–204. <https://doi.org/10.1080/17538157.2021.1879810>
- Torab-Miandoab, A., Basiri, M., Dabbagh-Moghaddam, A., & Gholamhosseini, L. (2025). Electronic Health Record in Military Healthcare Systems: A Systematic Review. *PLoS ONE*, 20(2), 1–24. <https://doi.org/10.1371/journal.pone.0313641>
- Torab-Miandoab, A., Samad-Soltani, T., Jodati, A., & Rezaei-Hachesu, P. (2023). Interoperability of Heterogeneous Health Information Systems: A Systematic Literature Review. *BMC Medical Informatics and Decision Making*, 23(1), 1–13. <https://doi.org/10.1186/s12911-023-02115-5>
- V.H., C., A.I., M.-G., & J.R., P. (2010). A Knowledge-Based Taxonomy of Critical Factors for Adopting Electronic Health Record Systems by Physicians: A Systematic Literature Review. *BMC Medical Informatics and Decision Making*, 10, 60. Retrieved from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emedx&NEWS=N&AN=20950458%5Chttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed9&NEWS=N&AN=20950458>
- Wang, D., & Zhang, S. (2024). Large Language Models in Medical and Healthcare Fields: Applications, Advances, and Challenges. *Artificial Intelligence Review*, 57, Article 10921. <https://doi.org/10.1007/s10462-024-10921-0>
- Winiasri, L., Santosa, T. A., Yohandri, Y., Razak, A., Festiyed, F., & Zulyusri, Z. (2023). Ethno-Biology Learning Model Based on Design Thinking to Improve Students' Critical Thinking Skills. *Jurnal Penelitian Pendidikan IPA*, 9(9), 7767–7774. <https://doi.org/10.29303/jppipa.v9i9.4213>
- Xanthidou, O. K., Shuib, L., Xanthidis, D., & Nicholas, D. (2018). Electronic Medical Records in Greece and Oman: A Professional's Evaluation of Structure and Value. *International Journal of Environmental Research and Public Health*, 15(6), 1137. <https://doi.org/10.3390/ijerph15061137>
- Xiang, Y., Xu, J., Si, Y., Li, Z., Rasmy, L., Zhou, Y., ... Wang, F. (2019). Time-Sensitive Clinical Concept Embeddings Learned from Large Electronic Health Records. *BMC Medical Informatics and Decision Making*, 19(Suppl 2), Article 58. <https://doi.org/10.1186/s12911-019-0766-3>
- Xiao, C., Choi, E., & Sun, J. (2018). Opportunities and Challenges in Developing Deep Learning Models Using Electronic Health Records Data: A Systematic Review. *Journal of the American Medical Association*. Retrieved from <https://academic.oup.com/jamia/article-abstract/25/10/1419/5035024>
- Zang, C., Hou, Y., Lyu, D., Jin, J., Sacco, S., Chen, K., ... Wang, X. (2024). Accuracy and Transportability of Machine Learning Models for Adolescent Suicide Prediction with Longitudinal Clinical Records. *Translational Psychiatry*, 14, Article 3034.

- <https://doi.org/10.1038/s41398-024-03034-3>
Zarour, M., Alenezi, M., Ansari, M. T. J., Pandey, A. K., Ahmad, M., Agrawal, A., Kumar, R., & Khan, R. A. (2021). Ensuring Data Integrity of Healthcare Information in the Era of Digital Health. *Healthcare Technology Letters*, 8(3), 66–77. <https://doi.org/10.1049/htl2.12008>
- Zhang, P., Shen, S., Deng, W., Mao, S., & Wang, Y. (2022). The Construction Model of the TCM Clinical Knowledge Coding Database Based on Knowledge Organization. *BioMed Research International*, 2022, Article 2503779. <https://doi.org/10.1155/2022/2503779>
- Zulkifli, Z., Satria, E., Supriyadi, A., & Santosa, T. A. (2022). Meta-Analysis: The Effectiveness of the Integrated STEM Technology Pedagogical Content Knowledge Learning Model on the 21st Century Skills of High School Students in the Science Department. *Psychology, Evaluation, and Technology in Educational Research*, 5(1), 32–42. <https://doi.org/10.33292/petier.v5i1.144>
- Zulyusri, Z., Santosa, T. A., Festiyed, F., Yerimadesi, Y., Yohandri, Y., Razak, A., & Sofianora, A. (2023). Effectiveness of STEM Learning Based on Design Thinking in Improving Critical Thinking Skills in Science Learning: A Meta-Analysis. *Jurnal Penelitian Pendidikan IPA*, 9(6), 112–119. <https://doi.org/10.29303/jppipa.v9i6.3709>