
PEMANFAATAN ARTIFICIAL INTELLIGENCE DALAM MANAJEMEN INFORMASI KESEHATAN UNTUK Mendukung KUALITAS DATA KESEHATAN: SCOPING REVIEW

Utilization of Artificial Intelligence in Health Information Management to Support Health Data Quality: A Scoping Review

Donna Dwinita Adelia^{1*}, Donny Yunamawan Luddie Sunaryo², Rokhamah³, Nisa'i Daramita Supriyono⁴

^{1,2,3,4} Politeknik Kesehatan Wira Husada Nusantara Malang

ARTICLE INFO

Article history

Submitted date:
22-07-2026
Received date:
31-07-2026
Published date:
31-07-2026

Keywords:

artificial intelligence;
health data quality; health
information management;
electronic health records;
scoping review

ABSTRACT

Digital transformation has expanded the volume and complexity of data managed through electronic health records and health information systems, while incompleteness, inconsistency, duplication, coding errors, and poor standardization remain common. This scoping review mapped the use of artificial intelligence (AI) in health information management and examined whether reported outcomes directly measured data quality or only downstream technical performance. Reporting followed PRISMA-ScR. Publications from January 2021 to 18 July 2026 were searched in PubMed and supplementary publisher platforms. Two reviewers independently screened records and resolved disagreements by consensus. Ten sources were included and separated into three primary studies and seven secondary reviews. Direct primary evidence showed that K-nearest-neighbor imputation reduced missing values in a processed dataset and that large-language-model preprocessing improved clinical concept extraction (F1-score 0.40 to 0.61). These findings reflect improvements in dataset completeness after imputation and downstream extraction performance, not proof that missing patient information was actually collected or that clinical outcomes improved. Evidence for workflow, organizational, and real-world clinical impact remained limited. AI therefore shows potential as a data-quality support and quality-control tool, but implementation requires validation, human oversight, bias monitoring, privacy protection, and alignment with interoperability standards.

Kata kunci:

artificial intelligence;
kualitas data kesehatan;
manajemen informasi
kesehatan; rekam medis
elektronik; scoping review

ABSTRAK

Transformasi digital meningkatkan volume dan kompleksitas data yang dikelola melalui rekam medis elektronik dan sistem informasi kesehatan, sementara masalah ketidaklengkapan, inkonsistensi, duplikasi, kesalahan pengodean, dan rendahnya standarisasi masih sering ditemukan. Scoping review ini memetakan pemanfaatan artificial intelligence (AI) dalam manajemen informasi kesehatan serta menilai apakah luaran yang dilaporkan mengukur kualitas data secara langsung atau hanya performa teknis hilir. Pelaporan mengacu pada PRISMA-ScR. Publikasi Januari 2021 hingga 18 Juli 2026 ditelusuri melalui PubMed dan platform penerbit tambahan. Dua reviewer melakukan skrining secara independen dan menyelesaikan perbedaan melalui konsensus. Sepuluh sumber dimasukkan, terdiri atas tiga penelitian primer dan tujuh artikel review. Bukti primer langsung menunjukkan bahwa imputasi K-nearest neighbors mengurangi nilai kosong pada dataset yang telah diproses dan prapemrosesan large language model meningkatkan F1-score ekstraksi konsep klinis dari 0,40 menjadi 0,61. Temuan tersebut menggambarkan kelengkapan dataset setelah imputasi dan peningkatan performa ekstraksi, bukan bukti bahwa informasi pasien yang hilang benar-benar terkumpul atau bahwa luaran klinis membaik. Bukti mengenai dampak alur kerja, organisasi, dan praktik klinis nyata masih terbatas. AI menunjukkan potensi sebagai alat pendukung pengendalian kualitas data, tetapi penerapannya memerlukan validasi, pengawasan manusia, pemantauan bias, perlindungan privasi, dan kesesuaian dengan standar interoperabilitas.

Corresponding Author:

Donna Dwinita Adelia

PENDAHULUAN

Perkembangan teknologi digital telah mengubah proses pencatatan, penyimpanan, pengolahan, dan pertukaran informasi kesehatan. Di Indonesia, Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 menetapkan penyelenggaraan rekam medis elektronik pada fasilitas pelayanan kesehatan. Integrasi data nasional selanjutnya diarahkan melalui SATUSEHAT, yang menerapkan HL7 FHIR sebagai standar model data dan antarmuka pertukaran informasi serta menggunakan terminologi seperti ICD-10, ICD-9-CM, SNOMED CT, dan LOINC (Kementerian Kesehatan Republik Indonesia, 2022; SATUSEHAT, 2026a, 2026b).

Peningkatan volume data elektronik tidak otomatis menjamin kualitas data. Rekam medis dapat memuat nilai yang hilang, format tidak sesuai, diagnosis tidak terkode, duplikasi identitas, terminologi tidak konsisten, kesalahan pengetikan, nilai tidak masuk akal, dan keterlambatan pembaruan. Literatur kualitas data kesehatan digital secara konsisten menempatkan kelengkapan, akurasi atau kebenaran, konsistensi atau kesesuaian, kewajaran, dan kemutakhiran sebagai dimensi penting untuk penggunaan data klinis dan sekunder (Lewis et al., 2023; Penev et al., 2024; Syed et al., 2023).

Artificial intelligence (AI) menawarkan berbagai pendekatan untuk mendukung pemeriksaan kualitas data dalam skala besar. Machine learning dapat digunakan untuk mendeteksi anomali dan pola kesalahan, natural language processing (NLP) dapat mengubah narasi klinis menjadi elemen terstruktur, sedangkan large language model (LLM) dapat membantu koreksi bahasa dan standarisasi istilah. Namun, luaran seperti akurasi klasifikasi, F1-score, atau AUC tidak selalu merupakan ukuran langsung kualitas data. Luaran tersebut dapat merefleksikan performa model hilir setelah data diproses, sehingga perlu dibedakan dari perubahan kelengkapan, akurasi, validitas, konsistensi, keunikan, ketepatan waktu, atau standarisasi data itu sendiri.

Selain itu, performa AI dipengaruhi oleh kualitas dan representativitas data pelatihan, variasi sistem rekam medis, ketidakseimbangan kelas, kemampuan generalisasi, transparansi algoritma, keamanan informasi, serta keterlibatan manusia. WHO menekankan bahwa penggunaan AI untuk kesehatan memerlukan tata kelola yang menjamin kualitas data, keadilan, keamanan, akuntabilitas, dan pengawasan manusia (World Health Organization, 2021; World Health Organization Regional Office for Europe, 2025).

Karena sumber yang tersedia mencakup penelitian primer, systematic review, dan scoping review dengan outcome yang heterogen, kajian ini direvisi menjadi scoping review. Tujuannya adalah memetakan bentuk pemanfaatan AI dalam manajemen informasi kesehatan, mengelompokkan outcome kualitas data langsung dan tidak langsung, membedakan tingkat bukti teknis, proses kerja, serta organisasi atau klinis, dan mengidentifikasi kesenjangan implementasi.

METODE PENELITIAN

Desain kajian

Penelitian ini menggunakan desain scoping review karena pertanyaan kajian bersifat luas dan sumber bukti yang dipetakan mencakup penelitian primer maupun artikel review. Pelaporan mengikuti Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR), sedangkan alur identifikasi dan seleksi dilaporkan menggunakan diagram PRISMA (Tricco et al., 2018; Page et al., 2021). Artikel review diperlakukan sebagai bukti sekunder dan tidak digunakan untuk menghitung ulang efek dari penelitian primer yang juga muncul dalam review tersebut.

Pertanyaan kajian

1. Teknologi AI apa yang digunakan dalam manajemen informasi kesehatan dan pengendalian kualitas data?
2. Dimensi kualitas data apa yang diukur secara langsung, dan outcome apa yang hanya menjadi bukti tidak langsung?
3. Pada tingkat mana bukti tersedia: performa algoritma, perbaikan proses kerja, atau dampak organisasi dan klinis?
4. Apa keterbatasan metodologis, risiko implementasi, dan kebutuhan tata kelola yang dilaporkan?

Kerangka PCC

Tabel 1. Kerangka Population–Concept–Context (PCC)

Komponen	Kriteria
Population	Data kesehatan, rekam medis elektronik, catatan klinis, registri, dan sistem pertukaran informasi kesehatan.
Concept	AI, machine learning, deep learning, NLP, anomaly detection, imputasi, automated coding, dan LLM yang berkaitan dengan kelengkapan, akurasi, konsistensi, validitas, keunikan, ketepatan waktu, standardisasi, interoperabilitas, atau efisiensi pengelolaan data.
Context	Manajemen informasi kesehatan, pelayanan kesehatan, penelitian berbasis data kesehatan, dan sistem informasi kesehatan.

Sumber informasi dan strategi penelusuran

Penelusuran terakhir dilakukan pada 18 Juli 2026 untuk publikasi Januari 2021–18 Juli 2026. PubMed digunakan sebagai basis data bibliografis utama. PubMed Central, ScienceDirect, SpringerLink, Frontiers, dan JMIR digunakan sebagai sumber tambahan untuk menemukan teks lengkap dan artikel pada platform penerbit. Seluruh hasil diimpor ke Zotero, kemudian duplikasi dikenali berdasarkan DOI, PMID, judul, penulis pertama, dan tahun. Pemeriksaan manual dilakukan untuk rekaman dengan metadata yang tidak lengkap. Tumpang tindih antara PubMed dan PubMed Central diperiksa secara khusus.

Tabel 2. Strategi pencarian per sumber

Sumber	String pencarian ringkas	Filter	Jumlah hasil
PubMed	("artificial intelligence"[Title/Abstract] OR "machine learning"[Title/Abstract] OR "deep learning"[Title/Abstract] OR "natural language processing"[Title/Abstract] OR "large language model"[Title/Abstract]) AND ("electronic health record"[Title/Abstract] OR "health information management"[Title/Abstract] OR "clinical documentation"[Title/Abstract] OR "health information exchange"[Title/Abstract]) AND ("data quality"[Title/Abstract] OR completeness[Title/Abstract] OR accuracy[Title/Abstract] OR consistency[Title/Abstract] OR validity[Title/Abstract] OR duplicate*[Title/Abstract] OR "clinical coding"[Title/Abstract])	2021/01/01–2026/07/18; Inggris/Indonesia	814
PubMed Central	Kombinasi istilah AI + EHR/clinical documentation + data quality/completeness/accuracy/consistency/validity/coding	2021–2026; artikel teks lengkap	213
ScienceDirect	TITLE-ABSTR-KEY: AI/ML/NLP/LLM AND EHR/health information AND data quality/completeness/accuracy/consistency/coding	2021–2026; research/review article	131
SpringerLink	AI/ML/NLP/LLM AND electronic health record/health information AND data quality/completeness/accuracy/consistency	2021–2026; article	100

Sumber	String pencarian ringkas	Filter	Jumlah hasil
Frontiers	AI/ML/NLP/LLM AND healthcare data quality/EHR quality	2021–2026; article	36
JMIR	AI/ML/NLP/LLM AND electronic health record/clinical notes AND data quality/completeness/coding	2021–2026; research/review	20

Sumber : Data Analysis, 2026

Kriteria inklusi dan eksklusi

Sumber dimasukkan apabila diterbitkan pada periode pencarian, berbahasa Inggris atau Indonesia, membahas pemanfaatan AI pada data atau informasi kesehatan, serta melaporkan outcome kualitas data langsung atau outcome teknis/proses yang dapat dipetakan sebagai bukti tidak langsung. Penelitian primer, systematic review, dan scoping review dapat dimasukkan, tetapi dianalisis dalam kelompok terpisah. Editorial, opini, berita, artikel komersial, studi citra medis tanpa hubungan dengan manajemen informasi, dataset nonkesehatan, artikel tanpa metode evaluasi, dan duplikat dikeluarkan. Studi yang hanya menilai prediksi atau deteksi tanpa mengukur dimensi kualitas data diklasifikasikan sebagai bukti tidak langsung, bukan bukti efektivitas kualitas data.

Seleksi sumber bukti

Dua reviewer melakukan skrining judul–abstrak dan pemeriksaan teks lengkap secara independen. Perbedaan keputusan diselesaikan melalui diskusi dan peninjauan ulang terhadap kriteria inklusi; keputusan akhir ditetapkan melalui konsensus kedua reviewer. Alasan eksklusi pada tahap teks lengkap dikelompokkan menjadi: outcome tidak terkait kualitas data, konteks bukan manajemen informasi kesehatan, desain atau metode tidak memenuhi kriteria, dan artikel duplikat atau versi publikasi yang sama.

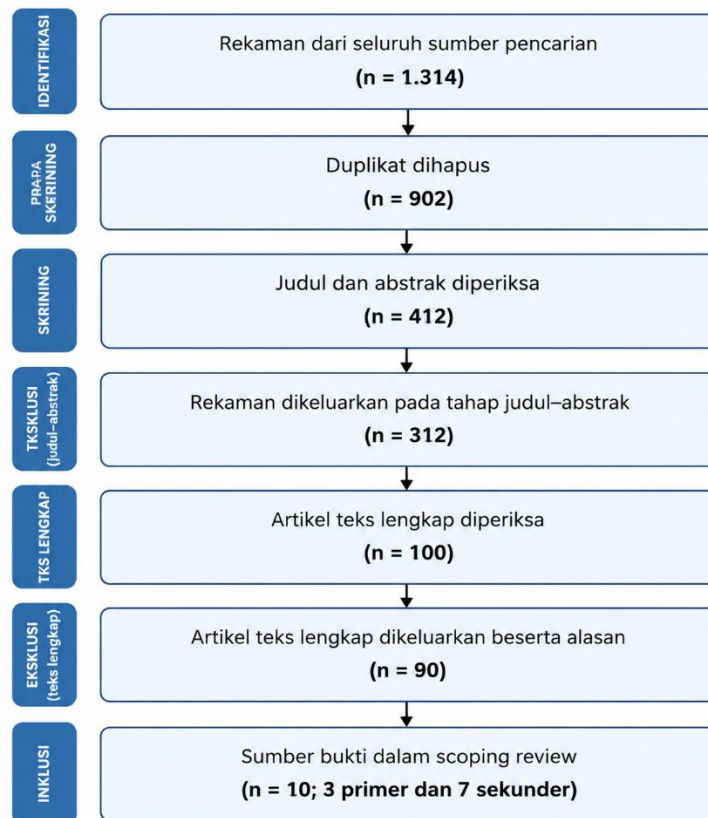
Ekstraksi, klasifikasi outcome, dan sintesis

Data yang diekstraksi meliputi penulis, tahun, desain, jenis data, teknologi AI, dimensi kualitas data, outcome utama, keterbatasan, dan tingkat bukti. Outcome langsung mencakup perubahan kelengkapan, akurasi, konsistensi, validitas, keunikan, ketepatan waktu, dan standardisasi. Outcome seperti F1-score klasifikasi, AUC prediksi, atau deteksi kejadian keselamatan ditempatkan sebagai bukti tidak langsung apabila tidak disertai pengukuran perubahan kualitas data sumber. Karena heterogenitas desain, dataset, algoritma, dan indikator, sintesis dilakukan secara naratif tanpa meta-analisis dan tanpa menentukan teknologi yang “paling efektif”.

Penilaian kualitas metodologis

Penelitian primer dinilai menggunakan Mixed Methods Appraisal Tool (MMAT) 2018 sesuai kategori desain, sedangkan artikel systematic review dinilai menggunakan AMSTAR 2. Untuk scoping review, unsur pelaporan PRISMA-ScR, transparansi strategi pencarian, proses seleksi, dan keterlacakan ekstraksi turut diperiksa. Penilaian digunakan untuk menjelaskan keterbatasan dan tidak digunakan sebagai skor pembobotan kuantitatif karena tujuan kajian adalah pemetaan bukti (Hong et al., 2018; Shea et al., 2017)

Gambar 1. Diagram alir seleksi sumber bukti



HASIL ANALISIS

Karakteristik dan pemisahan sumber bukti

Sebanyak 10 sumber bukti dipetakan, terdiri atas tiga penelitian primer dan tujuh artikel review. Pemisahan ini dilakukan untuk mencegah penghitungan bukti berulang. Penelitian primer digunakan untuk menggambarkan hasil empiris langsung pada dataset tertentu, sedangkan artikel review digunakan untuk memetakan tren metode, dimensi kualitas data, keterbatasan validasi, dan kesenjangan implementasi.

Tabel 3. Karakteristik sumber bukti dan keterkaitannya dengan kualitas data

Sumber	Jenis bukti	Teknologi/data	Temuan utama	Hubungan dengan kualitas data	Tingkat bukti
Fudickar et al., 2024	Primer, retrospektif; 1.608 pasien	NLP, SVM, KNN, MLP; surat rujukan	Informasi surat rujukan meningkatkan F1-score model triase hingga 19,5%, tetapi performa belum memadai untuk penggunaan mandiri.	Tidak langsung: performa prediksi setelah penambahan fitur, bukan perubahan dimensi kualitas data.	Performa algoritma

Sumber	Jenis bukti	Teknologi/data	Temuan utama	Hubungan dengan kualitas data	Tingkat bukti
Jarmakovic a, 2025	Primer, eksperimen; 768 rekaman diabetes	KNN imputation, anomaly detection, random forest, LightGBM	Proporsi nilai terisi meningkat dari 90,57% menjadi hampir 100%; random forest akurasi 75,3% dan AUC 0,83.	Langsung untuk kelengkapan dataset setelah imputasi; ketepatan nilai imputasi tetap mengandung ketidakpastian.	Performa teknis
Hier et al., 2025	Primer; 1.618 catatan neurologi	LLM dan NLP berbasis ontologi	Prapemrosesan meningkatkan F1-score ekstraksi HPO dari 0,40 menjadi 0,61; sampel ahli tidak menemukan kehilangan informasi bermakna.	Langsung untuk konsistensi/standarisasi teks; F1-score merupakan outcome hilir.	Performa teknis
Ozonze et al., 2023	Systematic review	Otomatisasi, NLP, ML, rule mining	Alat otomatis meningkat, tetapi pendekatan terfragmentasi dan standarisasi evaluasi lemah.	Bukti sekunder mengenai alat penilaian kualitas data.	Pemetaan metode
Lewis et al., 2023	Systematic review	Alat komputasional penilaian EHR	Kelengkapan paling sering dinilai; belum ada pendekatan standar yang universal.	Bukti sekunder langsung mengenai dimensi dan metode penilaian.	Pemetaan metode
Kaur et al., 2023	Systematic literature review	NLP/deep learning untuk pengodean ICD	Deep learning semakin banyak digunakan; korpus besar dan akurasi kode tetap menjadi kendala.	Sebagian langsung pada konsistensi/akurasi pengodean; sebagian berupa performa klasifikasi.	Pemetaan metode
Penev et al., 2024	Scoping review	Penilaian kualitas dan performa EHR	Kelengkapan, kesesuaian, kewajaran, dan akurasi dominan; indikator bergantung konteks penggunaan.	Bukti sekunder mengenai dimensi kualitas data.	Pemetaan konsep
Jarmakovic a, 2025 (review)	Review literatur	Supervised learning, NLP, anomaly detection, federated/data-centric AI	AI berpotensi mendukung akurasi dan konsistensi; timeliness dan validity jarang dievaluasi.	Bukti sekunder; tidak digunakan untuk menghitung ulang studi primer penulis yang sama.	Pemetaan metode

Sumber	Jenis bukti	Teknologi/data	Temuan utama	Hubungan dengan kualitas data	Tingkat bukti
Kayira et al., 2026	Scoping review; 226 studi	Rule-based NLP, BERT, transformer, model hibrida	25 studi melaporkan potensi cakupan/kelengkapan dan kecepatan ekstraksi; tidak ada evaluasi dampak klinis nyata.	Bukti sekunder; terutama performa teknis dan potensi proses.	Pemetaan implementasi
Deimazar & Sheikhtaheri, 2023	Systematic review; 41 studi	SVM, RF, CRF, RNN, BiLSTM-CRF	AI dapat mendeteksi kejadian keselamatan; mayoritas belum divalidasi eksternal atau diuji prospektif.	Tidak langsung terhadap kualitas data; fokus pada deteksi kejadian keselamatan.	Performa algoritma

Hasil penilaian kualitas metodologis

Penilaian menunjukkan bahwa seluruh sumber memiliki keterbatasan yang relevan terhadap interpretasi. Penelitian primer umumnya menggunakan data satu institusi atau satu dataset, sedangkan artikel review berbeda dalam kelengkapan pencarian, penilaian risiko bias, dan pelaporan daftar studi. Oleh karena itu, tidak ada sumber yang diperlakukan memiliki bobot bukti yang sama secara otomatis.

Tabel 4. Ringkasan penilaian kualitas

Sumber	Instrumen	Penilaian ringkas	Keterbatasan utama
Fudickar et al., 2024	MMAT	Sebagian besar kriteria terpenuhi	Retrospektif; outcome kualitas data tidak langsung; belum siap untuk penggunaan klinis mandiri.
Jarmakovica, 2025	MMAT	Sebagian besar kriteria terpenuhi	Dataset publik kecil; imputasi tidak membuktikan nilai pasien yang benar; validasi eksternal terbatas.
Hier et al., 2025	MMAT	Sebagian besar kriteria terpenuhi	Satu spesialisasi; ground truth sebagian dibantu LLM; tinjauan ahli hanya pada sampel.
Ozonze et al., 2023	AMSTAR 2	Kepercayaan sedang	Heterogenitas alat dan pelaporan kualitas studi primer terbatas.
Lewis et al., 2023	AMSTAR 2	Kepercayaan sedang	Tidak ada standar penilaian EHR yang universal; studi yang dimasukkan heterogen.
Kaur et al., 2023	AMSTAR 2	Kepercayaan sedang	Korpus, versi ICD, distribusi kode, dan metrik sangat heterogen.
Penev et al., 2024	PRISMA-ScR	Pelaporan memadai	Scoping review tidak ditujukan untuk estimasi efek; konteks penilaian beragam.
Jarmakovica, 2025 (review)	AMSTAR 2	Kepercayaan rendah-sedang	Bentuk review dan proses appraisal kurang rinci dibanding systematic review penuh.
Kayira et al., 2026	PRISMA-ScR	Pelaporan memadai	Sebagian besar studi satu institusi; tidak ada evaluasi dampak praktik klinis nyata.

Sumber	Instrumen	Penilaian ringkas	Keterbatasan utama
Deimazar & Sheikhtaheri, 2023	AMSTAR 2	Kepercayaan sedang	Kualitas rata-rata studi rendah; validasi eksternal dan pengujian prospektif jarang.

Pemetaan temuan berdasarkan dimensi kualitas data

Kelengkapan

Bukti primer langsung mengenai kelengkapan terutama berasal dari Jarmakovica (2025). Penggunaan KNN imputation mengurangi proporsi nilai kosong sehingga kelengkapan dataset setelah pemrosesan meningkat dari 90,57% menjadi hampir 100%. Angka tersebut tidak berarti bahwa informasi pasien yang sebelumnya hilang telah dikumpulkan kembali. Nilai baru merupakan estimasi algoritmik berdasarkan pola data lain dan tetap memiliki ketidakpastian. Karena itu, hasil imputasi perlu diuji melalui analisis sensitivitas, pemeriksaan distribusi, perbandingan dengan metode imputasi lain, serta validasi terhadap data rujukan bila tersedia.

NLP juga dapat memperluas cakupan data terstruktur dengan mengekstraksi informasi dari narasi klinis. Namun, tambahan elemen terstruktur baru dapat dianggap meningkatkan kelengkapan hanya jika informasi tersebut benar-benar terdapat dalam teks sumber dan hasil ekstraksi telah divalidasi. Kesalahan memahami negasi, singkatan, konteks temporal, atau hubungan antarkonsep tetap dapat menghasilkan data keliru.

Akurasi, konsistensi, dan standardisasi

Hier et al. (2025) menunjukkan bahwa prapemrosesan catatan klinis dengan LLM memperbaiki ejaan, tata bahasa, singkatan, istilah, dan format, kemudian meningkatkan F1-score ekstraksi konsep HPO dari 0,40 menjadi 0,61. Temuan ini mendukung potensi perbaikan konsistensi dan standardisasi teks serta peningkatan performa aplikasi hilir. Namun, evaluasi dilakukan pada satu spesialisasi dan validasi lebih luas tetap diperlukan untuk memastikan bahwa perubahan bahasa tidak menggeser makna klinis.

Temuan Fudickar et al. (2024) ditempatkan sebagai bukti tidak langsung. Peningkatan F1-score triase setelah informasi surat rujukan ditambahkan menunjukkan bahwa data tekstual tambahan dapat memperkaya fitur model, tetapi tidak secara langsung membuktikan peningkatan akurasi atau kelengkapan rekam medis. Demikian pula, deteksi kejadian keselamatan pasien dalam review Deimazar dan Sheikhtaheri (2023) berhubungan dengan pemanfaatan data EHR, tetapi outcome utamanya adalah performa deteksi, bukan kualitas data sumber.

Pengodean klinis, ketepatan waktu, keunikan, dan interoperabilitas

Review Kaur et al. (2023) menunjukkan bahwa NLP dan deep learning dapat mendukung rekomendasi kode ICD, tetapi kinerja dipengaruhi oleh ketidakseimbangan kode, variasi aturan pengodean, keterbatasan korpus, dan perubahan versi klasifikasi. Sistem tersebut lebih tepat digunakan untuk menyarankan kode atau memprioritaskan berkas untuk ditinjau daripada menggantikan keputusan petugas coding.

Bukti mengenai ketepatan waktu pembaruan data, pencegahan duplikasi identitas, dan interoperabilitas lintas institusi masih terbatas. AI dapat membantu pencocokan identitas dan pemetaan terminologi, tetapi keputusan penggabungan rekam medis dan transformasi elemen data harus melalui verifikasi manusia. Dalam konteks SATUSEHAT, keluaran AI harus mengikuti resource HL7 FHIR dan terminologi yang berlaku; interoperabilitas tidak dapat dicapai hanya melalui algoritma tanpa kesepakatan metadata, definisi variabel, identitas, keamanan, dan tata kelola (SATUSEHAT, 2026a, 2026b).

Tingkat kekuatan bukti

Tabel 5. Tingkat bukti pemanfaatan AI untuk kualitas data kesehatan

Tingkat	Contoh indikator	Ringkasan bukti	Interpretasi
Performa algoritma	Precision, recall, F1-score, AUC, deteksi anomali	Paling banyak dilaporkan; hasil beragam dan sering berasal dari satu institusi/dataset.	Menunjukkan kemampuan teknis pada kondisi pengujian, bukan efektivitas organisasi atau klinis.

Tingkat	Contoh indikator	Ringkasan bukti	Interpretasi
Perbaikan proses kerja	Waktu ekstraksi, pengurangan audit manual, cakupan data terstruktur	Dilaporkan sebagai potensi atau perbandingan terbatas; evaluasi implementasi rutin masih sedikit.	Menjanjikan, tetapi perlu studi sebelum–sesudah dan pengukuran beban kerja nyata.
Dampak organisasi/klinis	Penurunan kesalahan data, klaim lebih tepat, keselamatan pasien, hasil klinis	Bukti sangat terbatas; review Kayira et al. tidak menemukan evaluasi dampak klinis nyata.	Belum cukup untuk menyimpulkan efektivitas dalam praktik.

PEMBAHASAN

Interpretasi utama

Kajian ini menunjukkan bahwa AI memiliki potensi untuk mendukung kualitas data kesehatan pada tugas yang terukur dan dapat divalidasi, seperti imputasi nilai kosong, deteksi anomali, ekstraksi entitas, koreksi bahasa, standardisasi istilah, dan rekomendasi kode. Akan tetapi, kekuatan kesimpulan harus dibatasi pada konteks pengujian masing-masing studi. Heterogenitas algoritma, dataset, dan indikator tidak memungkinkan perbandingan langsung atau penentuan teknologi dan dimensi yang “paling efektif”.

Bukti terkuat masih berada pada tingkat performa teknis. Peningkatan F1-score atau AUC penting untuk evaluasi model, tetapi belum otomatis membuktikan perbaikan kualitas data yang bermakna bagi pelayanan. Untuk menunjukkan manfaat proses kerja, penelitian perlu mengukur waktu penyelesaian, jumlah koreksi manusia, beban audit, dan tingkat penerimaan pengguna. Untuk menunjukkan dampak organisasi atau klinis, penelitian perlu menilai perubahan kesalahan dokumentasi, ketepatan klaim, keselamatan pasien, keputusan klinis, dan keberlanjutan setelah implementasi.

Peran tenaga manajemen informasi kesehatan

AI tidak menggantikan kebutuhan tenaga manajemen informasi kesehatan. Peran tenaga tersebut justru bergeser ke pengelolaan definisi data, validasi keluaran, pengembangan aturan bisnis, audit informasi, pemantauan kinerja, dan penghubung antara kebutuhan klinis dengan sistem teknologi. Model human-in-the-loop paling sesuai untuk tugas berisiko tinggi, termasuk pengodean, koreksi informasi, penggabungan identitas pasien, dan pelepasan data.

Bias, validasi, dan generalisasi

Model AI dapat mempertahankan atau memperbesar kesalahan apabila data pelatihan tidak lengkap, tidak representatif, atau mencerminkan praktik dokumentasi yang bias. Evaluasi karena itu tidak cukup menggunakan akurasi keseluruhan; performa perlu diperiksa menurut fasilitas, wilayah, bahasa, umur, jenis kelamin, jenis layanan, dan kelompok relevan lainnya. Validasi internal, eksternal, prospektif, serta pemantauan setelah implementasi diperlukan untuk mendeteksi penurunan performa dan perubahan pola data.

Kayira et al. (2026) melaporkan bahwa 161 dari 226 penelitian NLP menggunakan dataset satu institusi dan tidak ada studi yang menilai dampak sistem dalam praktik klinis nyata. Deimazar dan Sheikhtaheri (2023) juga menunjukkan bahwa mayoritas sistem deteksi kejadian keselamatan belum melalui validasi eksternal atau pengujian prospektif. Temuan tersebut mendukung penggunaan bahasa “menunjukkan potensi” dan bukan “terbukti efektif” untuk implementasi layanan rutin.

Privasi, keamanan, dan tata kelola

Penggunaan AI pada data kesehatan meningkatkan kebutuhan terhadap pembatasan tujuan penggunaan, pengendalian akses, deidentifikasi, pencatatan aktivitas sistem, evaluasi penyedia teknologi, keamanan pertukaran data, dan mekanisme pertanggungjawaban. WHO menempatkan tata kelola data sebagai dasar sistem kesehatan digital yang dapat dipercaya dan menekankan kualitas,

representativitas, keadilan, privasi, transparansi, serta pengawasan manusia (World Health Organization, 2021; World Health Organization Regional Office for Europe, 2025).

Implikasi bagi fasilitas pelayanan kesehatan di Indonesia

Penerapan awal dapat diprioritaskan pada tugas berisiko relatif rendah dan mudah diaudit, seperti pemeriksaan format, penandaan nilai kosong, deteksi kemungkinan anomali, ekstraksi elemen data dari narasi, standarisasi istilah, dan penentuan prioritas berkas untuk audit. Penggunaan pada pengodean, penggabungan identitas pasien, atau keputusan klinis memerlukan kontrol yang lebih ketat. Semua keluaran yang dikirimkan ke SATUSEHAT harus mengikuti struktur FHIR dan terminologi yang berlaku serta lolos validasi sistem (SATUSEHAT, 2026a, 2026b).

Evaluasi implementasi perlu menggabungkan indikator teknis dan operasional: sensitivitas, spesifisitas, precision, recall, F1-score, tingkat kesalahan, proporsi koreksi manusia, waktu penyelesaian, penerimaan pengguna, perubahan kualitas data sebelum–sesudah, dampak pada kelompok berbeda, dan insiden keamanan. Audit berkala diperlukan agar model tidak digunakan di luar konteks validasinya.

Keterbatasan kajian

Kajian ini memiliki beberapa keterbatasan. Pertama, sumber pencarian mencakup satu basis data bibliografis utama dan beberapa platform penerbit; basis data teknologi seperti IEEE Xplore belum dimasukkan sehingga artikel relevan mungkin terlewat. Kedua, angka hasil per sumber dan rincian beberapa tahap seleksi belum tersedia dalam log yang dilampirkan dan harus dilengkapi sebelum pengiriman ulang. Ketiga, bukti primer hanya sedikit dan sebagian sumber merupakan review, sehingga keduanya dipisahkan untuk mencegah penghitungan berulang. Keempat, outcome dan metode sangat heterogen sehingga meta-analisis dan perbandingan efektivitas tidak dilakukan. Kelima, sebagian besar bukti berfokus pada performa teknis, menggunakan satu institusi atau dataset, dan belum mengevaluasi dampak organisasi atau klinis nyata.

KESIMPULAN

Artificial intelligence menunjukkan potensi untuk mendukung kualitas data dalam manajemen informasi kesehatan, terutama melalui pengurangan nilai kosong dengan imputasi, deteksi anomali, ekstraksi informasi dari narasi, koreksi dan standarisasi catatan klinis, serta rekomendasi pengodean. Namun, kelengkapan setelah imputasi tidak sama dengan keberhasilan mengumpulkan informasi pasien yang sebenarnya, dan peningkatan F1-score atau AUC terutama menggambarkan performa teknis aplikasi hilir.

Bukti paling kuat masih berada pada tingkat performa algoritma. Bukti mengenai perbaikan proses kerja rutin, ketepatan waktu, validitas, keunikan identitas, interoperabilitas lintas institusi, dan dampak organisasi atau klinis masih terbatas. Karena itu, belum tersedia dasar untuk menyatakan AI “paling efektif” atau telah terbukti efektif secara umum dalam praktik pelayanan kesehatan.

AI sebaiknya digunakan sebagai alat pendukung pengendalian kualitas dengan mekanisme human-in-the-loop. Implementasi memerlukan validasi lokal dan eksternal, pemantauan bias dan kinerja, perlindungan privasi, tata kelola yang jelas, kompetensi tenaga, serta kesesuaian dengan standar interoperabilitas dan terminologi SATUSEHAT.

DAFTAR PUSTAKA

- Agate Jarmakoviča. (2025). Artificial intelligence methods and approaches to improve data quality in healthcare data. *Artificial Intelligence in the Life Sciences*, 8, 100135.
- Deimazar, G., & Sheikhtaheri, A. (2023). Machine learning models to detect and predict patient safety events using electronic health records: A systematic review. *International Journal of Medical Informatics*, 180, 105246. <https://doi.org/10.1016/j.ijmedinf.2023.105246>
- Fudickar, S., Bantel, C., Spieker, J., Töpfer, H., Stegeman, P., Schiphorst Preuper, H. R., Reneman, M. F., Wolff, A. P., & Soer, R. (2024). Natural language processing of referral letters for machine learning-based triaging of patients with low back pain to the most appropriate intervention: Retrospective study. *Journal of Medical Internet Research*, 26, e46857. <https://doi.org/10.2196/46857>

- Hier, D. B., Carrithers, M. A., Platt, S. K., Nguyen, A., Giannopoulos, I., & Obafemi-Ajayi, T. (2025). Preprocessing of physician notes by LLMs improves clinical concept extraction without information loss. *Information*, 16(6), 446. <https://doi.org/10.3390/info16060446>
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O’Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). Mixed Methods Appraisal Tool (MMAT), version 2018: User guide. McGill University.
- Jarmakovica, A. J. A. (2025). Machine learning-based strategies for improving healthcare data quality: An evaluation of accuracy, completeness, and reusability. *Frontiers in Artificial Intelligence*, 8, 1621514. <https://doi.org/10.3389/frai.2025.1621514>
- Kaur, R., Ginige, J. A., & Obst, O. (2023). AI-based ICD coding and classification approaches using discharge summaries: A systematic literature review. *Expert Systems with Applications*, 213, 118997. <https://doi.org/10.1016/j.eswa.2022.118997>
- Kayira, A. B., Elyazori, H. R. A., Lybarger, K., Walter, F. M., Chelala, C., & Funston, G. (2026). Natural language processing of clinical notes for cancer research and patient care prior to widespread adoption of generative AI: Scoping review. *JMIR AI*, 5, e73481. <https://doi.org/10.2196/73481>
- Kementerian Kesehatan Republik Indonesia. (2022). Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 tentang Rekam Medis.
- 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.
- Ozonze, O., Scott, P. J., & Hopgood, A. A. (2023). Automating electronic health record data quality assessment. *Journal of Medical Systems*, 47, 23. <https://doi.org/10.1007/s10916-022-01892-2>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Penev, Y. P., Buchanan, T. R., Ruppert, M. M., Liu, M., Shekouhi, R., Guan, Z., et al. (2024). Electronic health record data quality and performance assessments: Scoping review. *JMIR Medical Informatics*, 12, e58130. <https://doi.org/10.2196/58130>
- SATUSEHAT. (2026a). FHIR: Standar data model dan API SATUSEHAT. Kementerian Kesehatan Republik Indonesia.
- SATUSEHAT. (2026b). Terminologi dan standar interoperabilitas SATUSEHAT. Kementerian Kesehatan Republik Indonesia.
- Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., et al. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358, j4008. <https://doi.org/10.1136/bmj.j4008>
- Syed, R., Eden, R., Makasi, T., Chukwudi, I., Mamudu, A., Kamalpour, M., et al. (2023). Digital health data quality issues: Systematic review. *Journal of Medical Internet Research*, 25, e42615. <https://doi.org/10.2196/42615>
- Tricco, A. C., Lillie, E., Zarin, W., O’Brien, K. K., Colquhoun, H., Levac, D., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- World Health Organization. (2021). Ethics and governance of artificial intelligence for health. World Health Organization.
- World Health Organization Regional Office for Europe. (2025). Health data governance in the age of artificial intelligence: Policy imperatives for the WHO European Region. World Health Organization.