

Artikel Penelitian

Diagnosis Penyakit Menggunakan Artificial Intelligence (AI): Konsep, Bukti Ilmiah, dan Implikasi Klinis

Disease Diagnosis Using Artificial Intelligence (AI): Concepts, Scientific Evidence, and Clinical Implications

Fajar Prasetya^{1*}, Edward^{2*}

¹Faculty of Pharmacy, University of Mulawarman, Samarinda, Indonesia

²Atma Jaya Teaching & Research Hospital, Jakarta, Indonesia

*Email korespondensi: fajarprasetya@farmasi.unmul.ac.id ; edward@atmajaya.id

Abstrak

Kecerdasan buatan (artificial intelligence/AI) telah berkembang menjadi teknologi transformatif dalam bidang diagnostik medis, terutama melalui kemampuannya menganalisis data klinis yang kompleks secara otomatis untuk mendukung deteksi penyakit, klasifikasi diagnosis, dan stratifikasi risiko. Kemajuan pesat dalam machine learning (ML) dan deep learning (DL) telah mendorong terobosan signifikan pada berbagai modalitas diagnostik, termasuk pencitraan medis, biosinyal, patologi, genomik, serta electronic health records (EHR). Artikel tinjauan ini mengkaji secara kritis peran AI dalam mendiagnosis masalah kesehatan dan penyakit, dengan fokus pada landasan metodologis, aplikasi klinis, evaluasi kinerja, serta tantangan translasi ke praktik klinik. Berbagai bukti dari studi-studi penting menunjukkan bahwa sistem AI mampu mencapai tingkat akurasi diagnostik yang sebanding atau bahkan melampaui tenaga kesehatan profesional pada tugas-tugas spesifik dan terdefinisi dengan baik, seperti skrining retinopati diabetik, deteksi kanker payudara, klasifikasi aritmia, dan skrining kanker paru. Namun demikian, meskipun hasil awal tampak menjanjikan, adopsi klinis secara luas masih menghadapi berbagai hambatan, antara lain bias dataset, keterbatasan validasi eksternal, kurangnya uji klinis prospektif, isu keterjelasan atau explainability model, serta ketidakpastian regulasi. Artikel ini mensintesis bukti ilmiah terkini, menyoroti standar metodologis dalam evaluasi AI diagnostik, serta membahas implikasi etis, hukum, dan klinis dari penerapan teknologi ini. AI sebaiknya dipahami bukan sebagai pengganti klinisi, melainkan sebagai alat pendukung keputusan yang dapat meningkatkan akurasi, konsistensi, dan efisiensi diagnosis apabila diintegrasikan secara bertanggung jawab ke dalam sistem pelayanan kesehatan.

Diterima: 04 Maret

2025

Disetujui: 10 April

2025

Publikasi : 31 Mei

2025

Copyright :

© 2025, Jurnal Sains dan Kesehatan (J. Sains.Kes.).

Published by Faculty of Pharmacy, University of Mulawarman, Samarinda, Indonesia. This is an Open Access article under the CC-BY-NC License



Kata kunci: kecerdasan buatan; diagnosis penyakit; pembelajaran mesin; pembelajaran mendalam; pendukung keputusan klinis; pencitraan medis; rekam medis elektronik

Abstract

Artificial intelligence (AI) has evolved into a transformative technology in medical diagnostics, particularly through its ability to automatically analyze complex clinical data to support disease detection, diagnostic classification, and risk stratification. Rapid advances in machine learning (ML) and deep learning (DL) have driven significant breakthroughs across various

diagnostic modalities, including medical imaging, biosignals, pathology, genomics, and electronic health records (EHRs). This review article critically examines the role of AI in diagnosing health problems and diseases, with a focus on methodological foundations, clinical applications, performance evaluation, and the challenges of translating AI innovations into clinical practice. Evidence from key studies demonstrates that AI systems can achieve diagnostic accuracy comparable to, or even surpassing, that of healthcare professionals in specific and well-defined tasks, such as diabetic retinopathy screening, breast cancer detection, arrhythmia classification, and lung cancer screening. Nevertheless, despite these promising early results, broad clinical adoption still faces multiple barriers, including dataset bias, limited external validation, lack of prospective clinical trials, issues of model interpretability or explainability, and regulatory uncertainties. This article synthesizes current scientific evidence, highlights methodological standards in the evaluation of diagnostic AI, and discusses the ethical, legal, and clinical implications of implementing these technologies. AI should be understood not as a replacement for clinicians, but rather as a decision-support tool that can enhance diagnostic accuracy, consistency, and efficiency when responsibly integrated into healthcare systems.

Keywords: *artificial intelligence; disease diagnosis; machine learning; deep learning; clinical decision support; medical imaging; electronic medical records*

1 Pendahuluan

Diagnosis yang akurat merupakan fondasi utama dalam pelayanan kesehatan yang efektif. Kesalahan diagnosis yang mencakup diagnosis yang terlewat, tertunda, atau keliru merupakan salah satu sumber utama bahaya yang dapat dicegah di seluruh dunia. Berbagai estimasi menunjukkan bahwa sekitar 5–10% pasien dewasa di layanan rawat jalan mengalami kesalahan diagnosis setiap tahunnya, dengan konsekuensi berupa morbiditas, mortalitas, dan beban ekonomi yang signifikan [1].

Kompleksitas manifestasi penyakit, meningkatnya multimorbiditas, fragmentasi alur pelayanan, serta ledakan volume data klinis secara kolektif telah memperberat beban kognitif yang harus ditanggung oleh tenaga kesehatan. Proses diagnosis tradisional sangat bergantung pada kemampuan manusia dalam mengenali pola, melakukan penalaran klinis, dan memanfaatkan pengalaman klinik sebelumnya. Meskipun klinisi berpengalaman mampu memberikan penilaian yang benuansa, kinerja manusia tetap memiliki keterbatasan, termasuk kelelahan, bias kognitif, serta information overload.

Kondisi inilah yang mendorong meningkatnya minat terhadap pendekatan komputasional yang mampu menganalisis data besar dan heterogen secara cepat dan konsisten. Dalam konteks tersebut, kecerdasan buatan muncul sebagai alat potensial untuk mendukung pengambilan keputusan diagnostik dengan cara memperkuat kemampuan kognitif manusia, bukan menggantikannya [2].

Penerapan AI dalam dunia medis sebenarnya bukanlah fenomena baru. Sistem pakar awal pada era 1970–1980-an, seperti MYCIN untuk penatalaksanaan penyakit infeksi, menggunakan logika berbasis aturan dan pengetahuan yang dikodekan secara manual. Meskipun inovatif pada masanya, sistem-sistem ini menghadapi keterbatasan dalam skalabilitas, pemeliharaan, dan generalisasi, sehingga dampak klinis jangka panjangnya relatif terbatas.

Kebangkitan kembali AI di sektor kesehatan didorong terutama oleh kemajuan dalam machine learning, khususnya deep learning, yang didukung oleh pertumbuhan eksponensial data kesehatan digital dan peningkatan kapasitas komputasi [3]. Berbeda dengan sistem berbasis aturan, model AI modern mampu mempelajari pola langsung dari data, sehingga dapat menangani hubungan nonlinier dan kompleks yang menjadi ciri khas fenomena biologis dan klinis. Beberapa tonggak penting yang mempercepat perkembangan AI diagnostik meliputi digitalisasi pencitraan medis dan preparat patologi, meluasnya penggunaan EHR dan biosensor wearable, ketersediaan dataset besar yang teranotasi, serta terobosan dalam arsitektur jaringan saraf. Kombinasi faktor-faktor ini telah menggeser AI dari sekadar janji teoretis menjadi alat diagnostik yang memiliki nilai praktis di berbagai bidang kedokteran.

Dalam konteks diagnostik, AI merujuk pada sistem komputasional yang dirancang untuk melakukan tugas-tugas yang secara tradisional membutuhkan kecerdasan manusia, seperti mengenali pola penyakit atau membuat inferensi diagnostik. Di dalam payung AI, machine learning mencakup algoritma yang belajar dari data untuk meningkatkan kinerjanya seiring waktu tanpa pemrograman eksplisit. Deep learning merupakan subbidang dari ML yang menggunakan jaringan saraf tiruan berlapis banyak, sehingga mampu mengekstraksi fitur secara hierarkis dari data mentah. Pendekatan ini sangat efektif dalam diagnostik medis karena banyak input klinis seperti citra, sinyal fisiologis, dan teks klinis bersifat berdimensi tinggi dan tidak terstruktur.

Dalam praktik diagnostik, beberapa paradigma pembelajaran yang relevan meliputi pembelajaran terawasi (menggunakan data berlabel), pembelajaran tak terawasi (menemukan pola tanpa label), serta pembelajaran self-supervised atau weakly supervised yang memanfaatkan label parsial atau label proksi. Hingga saat ini, sebagian besar sistem AI diagnostik yang dikembangkan atau diimplementasikan masih mengandalkan pendekatan pembelajaran terawasi atau weakly supervised [4]. Diagnosis penyakit berbasis AI tidak selalu berarti pengambilan keputusan klinis secara otonom. Sebaliknya, sistem AI dapat berperan dalam berbagai tahap alur kerja klinis, seperti skrining dan triase dengan mengidentifikasi kasus normal atau memprioritaskan pasien berisiko tinggi, deteksi dan klasifikasi penyakit, penilaian tingkat keparahan atau stadium penyakit, serta dukungan keputusan diagnostik melalui keluaran probabilistik atau daftar diagnosis banding.

Selain itu, AI juga banyak digunakan dalam prediksi risiko kejadian penyakit yang akan segera terjadi, seperti perburukan sepsis, yang dalam praktiknya sering berfungsi sebagai alat diagnostik dini [5]. Oleh karena itu, nilai klinis AI sangat bergantung pada tujuan penggunaannya, populasi sasaran, serta posisinya dalam jalur diagnosis. Penelitian AI diagnostik banyak difokuskan pada bidang-bidang dengan ketersediaan data yang melimpah, standar pelabelan yang jelas, dan titik akhir diagnosis yang terdefinisi dengan baik. Area utama meliputi pencitraan medis (radiologi, oftalmologi, dermatologi), patologi komputasional berbasis whole-slide imaging, analisis biosinyal seperti EKG dan EEG, pemanfaatan EHR baik terstruktur maupun tidak terstruktur, serta diagnostik genomik dan multi-omik. Di antara berbagai bidang tersebut, diagnostik berbasis citra berkembang paling pesat karena tersedianya dataset digital berskala besar dan tolok ukur kinerja yang jelas dibandingkan pembaca ahli [6].

Meskipun banyak sistem AI melaporkan tingkat akurasi diagnostik yang tinggi, seperti nilai AUROC di atas 0,90, kinerja teknis semata tidak cukup untuk membenarkan penerapan klinis. Kegunaan diagnostik yang sesungguhnya bergantung pada validitas eksternal lintas populasi, kalibrasi dan reliabilitas model, integrasi dengan alur kerja klinis, dampaknya terhadap luaran pasien, serta kepatuhan terhadap aspek etis, hukum, dan regulasi. Berbagai tinjauan sistematis menunjukkan bahwa sebagian besar studi AI diagnostik masih bersifat retrospektif, berpusat pada satu institusi, dan rentan terhadap bias, sehingga membatasi penerapan di dunia nyata [7,8]. Kesadaran akan hal ini mendorong pengembangan standar pelaporan dan evaluasi seperti CONSORT-AI, SPIRIT-AI, dan TRIPOD+AI guna meningkatkan transparansi dan reproduksibilitas penelitian AI diagnostik.

Artikel ini bertujuan menyajikan tinjauan komprehensif dan kritis mengenai diagnosis penyakit berbasis AI dengan fokus pada: landasan metodologis sistem AI diagnostik, sintesis bukti dari berbagai domain aplikasi klinis utama, evaluasi kekuatan dan keterbatasan penelitian AI diagnostik saat ini, serta pembahasan aspek etis, regulatori, dan translasi ke praktik klinik. Struktur artikel ini meliputi bagian metode yang menjelaskan desain tinjauan dan kerangka evaluasi, bagian hasil dan pembahasan yang mengulas aplikasi klinis, kinerja, bias, dan tata kelola, serta bagian kesimpulan yang merangkum arah masa depan dan rekomendasi untuk adopsi AI diagnostik yang bertanggung jawab.

2 Metode Penelitian

2.1 Desain dan pendekatan kajian

Artikel ini disusun menggunakan pendekatan tinjauan naratif integratif dengan prinsip-prinsip sistematis, yang difokuskan pada penerapan kecerdasan buatan (artificial intelligence/AI) dalam

diagnosis masalah kesehatan dan penyakit. Pendekatan ini dipilih karena penelitian AI diagnostik sangat heterogen dari sisi modalitas data, konteks klinis, serta metode evaluasi, sehingga tidak memungkinkan dilakukannya meta-analisis kuantitatif tunggal yang valid (7,24). Pendekatan naratif memungkinkan integrasi bukti lintas disiplin sekaligus analisis kritis terhadap implikasi klinis dan metodologis AI diagnostik, sebagaimana direkomendasikan dalam kajian sistematis mutakhir terkait AI di bidang medis (2,8).

2.2 Kerangka konseptual: AI sebagai intervensi diagnostik

Dalam kajian ini, sistem AI diposisikan sebagai intervensi diagnostik klinis, bukan sekadar alat teknis. Perspektif ini sejalan dengan pendekatan regulatori dan klinis modern yang memandang algoritme diagnostik sebagai teknologi medis yang dapat memengaruhi keputusan klinis dan keselamatan pasien (17,19). Evaluasi sistem AI dilakukan melalui lima dimensi utama, yaitu definisi tugas klinis, modalitas data, pendekatan algoritmik, metode evaluasi kinerja, serta implikasi etika dan regulasi, sebagaimana direkomendasikan dalam kerangka TRIPOD dan TRIPOD+AI (20,21).

2.3 Sumber data dan identifikasi literatur

Literatur diidentifikasi melalui penelusuran terstruktur pada basis data ilmiah utama seperti PubMed/MEDLINE, Scopus, Web of Science, serta jurnal bereputasi tinggi (Nature Medicine, npj Digital Medicine, BMJ, JAMA). Strategi pencarian mengombinasikan istilah artificial intelligence, machine learning, deep learning, dan diagnosis (5,6). Pemilihan literatur berfokus pada artikel yang telah melalui proses peer review dan banyak disitusi, guna memastikan validitas metodologis dan relevansi klinis (13).

2.4 Kriteria inklusi dan eksklusi

Studi dimasukkan apabila mengevaluasi AI untuk tujuan diagnosis atau inferensi diagnostik pada manusia, menggunakan data klinis nyata, melaporkan metrik kinerja kuantitatif, serta memiliki DOI yang dapat diverifikasi (7).

Studi dikecualikan apabila bersifat teknis murni tanpa konteks klinis, hanya berfokus pada optimasi terapi, menggunakan data sintetis, atau tidak memenuhi standar transparansi pelaporan sebagaimana dianjurkan dalam CONSORT-AI dan TRIPOD (18,20).

2.5 Klasifikasi modalitas diagnostik berbasis AI

Studi diklasifikasikan berdasarkan modalitas data utama, yaitu pencitraan medis, patologi komputasional, biosinyal, dan rekam medis elektronik. Pendekatan ini mengikuti rekomendasi survei sistematis AI medis yang menekankan perbedaan tantangan dan kesiapan klinis antar-modalitas (2,6).

2.6 Metrik evaluasi kinerja diagnostik

Evaluasi kinerja diagnostik mengacu pada metrik standar seperti sensitivitas, spesifitas, nilai prediktif, akurasi, dan AUROC. Selain itu, kalibrasi model dinilai sebagai aspek penting untuk pengambilan keputusan klinis yang aman (11,22). Kajian ini menekankan bahwa metrik diskriminatif tinggi tidak selalu berbanding lurus dengan manfaat klinis, terutama pada skrining penyakit dengan prevalensi rendah (8).

2.7 Strategi validasi dan generalisasi

Validasi diklasifikasikan menjadi internal, eksternal, dan prospektif. Validasi eksternal dan prospektif dianggap sebagai indikator utama kesiapan translasi klinis, sebagaimana ditegaskan dalam evaluasi sistematis AI diagnostik (7,11).

2.8 Penilaian risiko bias dan kualitas metodologis

Penilaian risiko bias dilakukan secara konseptual dengan mengacu pada PROBAST, TRIPOD, dan TRIPOD+AI, yang merupakan standar internasional dalam evaluasi model prediksi dan diagnosis

klinis (21,23). Aspek utama yang dianalisis meliputi bias seleksi, bias pelabelan, overfitting, dan pelaporan analisis subkelompok (7).

2.9 Pertimbangan etika dan keadilan algoritmik

Kajian ini menilai sejauh mana studi mempertimbangkan bias demografis dan ketimpangan struktural. Pendekatan ini merujuk pada bukti bahwa algoritme kesehatan dapat memperkuat ketidakadilan jika tidak dirancang secara adil (16).

2.10 Transparansi dan explainable AI

Explainable AI dinilai sebagai elemen pendukung kepercayaan klinisi dan tata kelola, meskipun tidak mengantikan validasi eksternal dan uji klinis (2,18).

2.11 Pendekatan sintesis dan pelaporan

Sintesis dilakukan secara tematik dan naratif untuk mencerminkan kompleksitas bukti AI diagnostik lintas modalitas, sebagaimana dianjurkan dalam kajian integratif AI medis (8).

3 Hasil dan Pembahasan

3.1 Gambaran umum temuan utama

Hasil kajian menunjukkan bahwa bukti terkuat AI diagnostik berasal dari domain berbasis citra dan sinyal, di mana deep learning menunjukkan performa tinggi dan konsisten dalam tugas-tugas diagnostik terdefinisi sempit (2,6). Namun, perbedaan antara akurasi teknis dan dampak klinis nyata masih menjadi isu utama dalam translasi AI ke praktik klinis (7).

3.2 Diagnostik berbasis AI pada pencitraan medis

3.2.1 Radiologi

Model deep learning mampu mendeteksi berbagai kelainan radiologis dengan kinerja tinggi, termasuk kanker paru, pneumonia, dan perdarahan intrakranial (12,13). Namun, performa tersebut sangat bergantung pada kesesuaian populasi data pelatihan dan konteks klinis, sehingga AI lebih tepat diposisikan sebagai alat triase dan pendukung keputusan (7).

3.2.2 Mammografi

Evaluasi internasional menunjukkan bahwa AI dapat mencapai performa setara radiolog dalam skrining kanker payudara (11). Meski demikian, risiko false positive dan implikasi sistem layanan kesehatan menuntut evaluasi klinis menyeluruh sebelum implementasi luas (8).

3.2.3 Oftalmologi

AI untuk skrining retinopati diabetik merupakan contoh paling matang dari translasi klinis AI diagnostik, dengan bukti uji klinis prospektif dan penerapan otonom di layanan primer (9,10). Keberhasilan ini dipengaruhi oleh kesesuaian antara penyakit, metode diagnosis, dan kebutuhan sistem kesehatan (2).

3.3 Patologi komputasional

Pendekatan weakly supervised deep learning memungkinkan analisis whole-slide imaging tanpa anotasi detail, dengan kinerja tinggi pada deteksi kanker (4).

Namun, tantangan variasi pewarnaan dan perbedaan pemindai menuntut validasi lintas laboratorium sebelum implementasi luas (6).

3.4 Diagnostik berbasis biosinyal

3.4.1 ECG

Deep learning menunjukkan kinerja setara ahli kardiologi dalam klasifikasi aritmia berbasis ECG satu sadapan (14). AI berpotensi meningkatkan deteksi aritmia jarang dan berisiko tinggi dalam pemantauan jangka panjang (2).

3.4.2 Wearable devices

AI memungkinkan analisis data biosinyal kontinu dari perangkat wearable untuk diagnosis dini penyakit kardiovaskular dan gangguan tidur (5). Namun, kualitas data, noise, dan isu privasi tetap menjadi tantangan utama implementasi klinis (15).

3.5 Perbandingan kinerja AI dan klinisi

AI dapat menyamai atau melampaui klinisi pada tugas diagnostik sempit, tetapi belum mampu menggantikan penalaran klinis holistik manusia (2,7). Pendekatan kolaboratif manusia–AI dianggap sebagai model paling aman dan realistik (2).

3.6 Sintesis awal

AI telah menunjukkan kematangan teknis signifikan dalam diagnosis berbasis citra dan sinyal, namun keberhasilan klinis bergantung pada validasi eksternal, integrasi alur kerja, dan tata kelola yang bertanggung jawab (9,17).

3.7 Diagnosis dan inferensi klinis berbasis rekam medis elektronik (EHR)

Rekam medis elektronik (Electronic Health Records/EHR) merepresentasikan sumber data klinis yang sangat kaya dan kompleks, mencakup informasi demografis, diagnosis, hasil laboratorium, catatan klinis bebas, serta data longitudinal perjalanan penyakit. Berbeda dengan pencitraan medis yang relatif terstandarisasi, EHR bersifat heterogen, tidak lengkap, dan sangat dipengaruhi oleh praktik dokumentasi klinis di masing-masing institusi (5).

Berbagai studi menunjukkan bahwa model deep learning mampu memanfaatkan data EHR untuk melakukan inferensi diagnostik dan prediksi kejadian klinis, seperti sepsis, gagal napas, atau perburuan kondisi pasien, dengan tingkat akurasi yang tinggi. Dalam praktik klinis, prediksi semacam ini sering kali berfungsi sebagai diagnosis dini atau peringatan klinis (early warning systems), meskipun secara terminologi sering dikategorikan sebagai risk prediction (15). Namun, pembahasan kritis menunjukkan bahwa performa AI berbasis EHR sangat bergantung pada kualitas data input, kelengkapan pencatatan, serta stabilitas sistem informasi kesehatan. Model yang dilatih pada satu rumah sakit sering kali mengalami penurunan kinerja ketika diterapkan di institusi lain, mencerminkan fenomena dataset shift yang signifikan (7).

3.8 Kesenjangan antara performa algoritmik dan manfaat klinis

Meskipun banyak model AI berbasis EHR melaporkan nilai AUROC yang tinggi, bukti menunjukkan bahwa performa diskriminatif tidak selalu berbanding lurus dengan peningkatan luaran pasien. Beberapa sistem peringatan dini gagal menunjukkan manfaat klinis karena integrasi alur kerja yang buruk, alert fatigue, atau ketidakpercayaan klinisi terhadap rekomendasi algoritme (2).

Kajian sistematis menegaskan bahwa sebagian besar studi AI diagnostik berbasis EHR masih bersifat retrospektif dan jarang dievaluasi melalui uji klinis prospektif terkontrol. Hal ini membatasi kemampuan untuk menyimpulkan dampak nyata AI terhadap keselamatan pasien dan efisiensi layanan kesehatan (7,8).

3.9 Bias algoritmik dan keadilan dalam diagnosis berbasis AI

Salah satu temuan paling penting dalam literatur AI kesehatan adalah adanya risiko bias algoritmik yang dapat memperburuk ketimpangan kesehatan. Studi landmark menunjukkan bahwa algoritme kesehatan yang menggunakan biaya layanan sebagai proksi kebutuhan kesehatan secara sistematis merugikan kelompok ras tertentu, meskipun model tersebut tampak "akurat" secara statistik (15).

Dalam konteks diagnosis penyakit, bias dapat muncul dari berbagai sumber, termasuk distribusi data yang tidak seimbang, perbedaan prevalensi penyakit antar kelompok, serta bias dalam proses

pelabelan data klinis. Jika tidak ditangani secara eksplisit, AI diagnostik berpotensi memperkuat bias struktural yang telah ada dalam sistem kesehatan (5).

Oleh karena itu, pembahasan menekankan pentingnya evaluasi kinerja model pada subkelompok demografis, transparansi dalam pemilihan variabel, serta penerapan strategi mitigasi bias sebagai bagian integral dari pengembangan dan validasi AI diagnostik (2).

3.10 Dataset shift dan tantangan generalisasi

Dataset shift perubahan distribusi data antara fase pelatihan dan penerapan merupakan tantangan utama dalam implementasi AI diagnostik. Perubahan populasi pasien, protokol klinis, perangkat medis, atau praktik dokumentasi dapat menyebabkan penurunan kinerja model secara signifikan (7).

Dalam pembahasan ini, dataset shift dipandang sebagai risiko keselamatan pasien yang nyata, bukan sekadar isu teknis. Oleh karena itu, validasi eksternal lintas institusi dan pemantauan kinerja pasca-deployment menjadi prasyarat penting sebelum dan setelah AI digunakan dalam diagnosis klinis (17).

3.11 Explainable AI dan kepercayaan klinisi

Explainable AI (XAI) sering dipromosikan sebagai solusi untuk meningkatkan transparansi dan kepercayaan klinisi terhadap sistem AI diagnostik. Metode seperti saliency maps, feature attribution, dan attention mechanisms memungkinkan visualisasi kontribusi fitur terhadap prediksi model (18).

Namun, kajian kritis menunjukkan bahwa banyak metode explainability bersifat heuristik dan tidak selalu mencerminkan proses pengambilan keputusan internal model secara akurat. Oleh karena itu, explainability tidak boleh dipandang sebagai pengganti validasi klinis dan uji prospektif, melainkan sebagai komponen tambahan dalam kerangka tata kelola AI medis (19).

3.12 Implikasi etika dan tanggung jawab klinis

Penerapan AI dalam diagnosis penyakit menimbulkan pertanyaan etis mendasar terkait tanggung jawab klinis, informed consent, dan akuntabilitas. Ketika AI berkontribusi terhadap keputusan diagnostik, batas tanggung jawab antara pengembang, institusi, dan tenaga medis menjadi kabur (17). Dari perspektif etika medis, AI diagnostik harus mematuhi prinsip beneficence, non-maleficence, autonomy, dan justice. Hal ini mencakup perlindungan data pasien, pencegahan diskriminasi algoritmik, serta keterlibatan manusia dalam keputusan klinis yang berdampak signifikan (15).

3.13 Regulasi dan tata kelola AI diagnostik

Regulasi AI dalam bidang kesehatan berkembang pesat seiring meningkatnya adopsi teknologi ini. Otoritas regulator menekankan perlunya pendekatan lifecycle regulation, di mana AI dipantau tidak hanya pada tahap persetujuan awal, tetapi juga sepanjang siklus penggunaannya (17).

Pedoman pelaporan seperti CONSORT-AI dan SPIRIT-AI dirancang untuk meningkatkan kualitas bukti klinis AI, sementara TRIPOD dan TRIPOD+AI bertujuan memperbaiki transparansi pelaporan model prediksi dan diagnosis (18,19,21). Pembahasan ini menegaskan bahwa tanpa kerangka regulasi dan tata kelola yang kuat, potensi AI diagnostik untuk meningkatkan kualitas layanan kesehatan dapat terhambat atau bahkan menimbulkan risiko baru.

3.14 Sintesis lanjutan

Secara keseluruhan, hasil dan pembahasan pada bagian ini menunjukkan bahwa tantangan utama AI diagnostik tidak lagi terletak pada kemampuan teknis semata, melainkan pada aspek generalisasi, keadilan, transparansi, dan tata kelola klinis. Keberhasilan implementasi AI dalam diagnosis penyakit sangat bergantung pada integrasi multidisipliner antara teknologi, klinik, etika, dan regulasi (2,8).

4 Kesimpulan

Kajian ini menunjukkan bahwa kecerdasan buatan (artificial intelligence/AI) telah mengalami perkembangan pesat sebagai teknologi pendukung diagnosis masalah kesehatan dan penyakit, terutama dalam domain yang berbasis pola data seperti pencitraan medis, patologi digital, biosinyal,

dan rekam medis elektronik. Bukti ilmiah dari berbagai studi menunjukkan bahwa model machine learning dan deep learning mampu mencapai kinerja diagnostik yang setara, dan dalam beberapa konteks bahkan melampaui, tenaga medis manusia pada tugas-tugas yang terdefinisi secara sempit dan terstandarisasi.

Namun demikian, hasil kajian ini juga menegaskan bahwa keberhasilan diagnostik berbasis AI tidak dapat diukur semata-mata dari akurasi algoritmik. Kesenjangan antara performa teknis dan manfaat klinis nyata masih menjadi tantangan utama dalam translasi AI ke praktik klinis sehari-hari. Banyak sistem AI yang menunjukkan nilai AUROC tinggi dalam studi retrospektif gagal memberikan dampak signifikan terhadap luaran pasien ketika diterapkan dalam alur kerja klinis yang kompleks dan dinamis.

Dari perspektif metodologis, kualitas bukti AI diagnostik saat ini masih bervariasi. Sebagian besar studi masih terbatas pada validasi internal, dengan sedikit yang melakukan validasi eksternal lintas institusi atau uji klinis prospektif. Risiko bias baik akibat ketidakseimbangan data, pelabelan yang tidak sempurna, maupun dataset shift merupakan isu krusial yang dapat memengaruhi keselamatan pasien dan keadilan layanan kesehatan jika tidak ditangani secara sistematis.

Kajian ini juga menyoroti bahwa bias algoritmik bukanlah sekadar masalah teknis, melainkan isu etika dan keadilan kesehatan. Tanpa evaluasi subkelompok dan strategi mitigasi bias, AI diagnostik berpotensi memperkuat ketimpangan struktural yang telah ada dalam sistem kesehatan. Oleh karena itu, keadilan algoritmik harus menjadi komponen inti dalam pengembangan, validasi, dan penerapan AI diagnostik.

Explainable AI (XAI) memiliki peran penting dalam meningkatkan transparansi dan kepercayaan klinisi, namun tidak dapat menggantikan kebutuhan akan validasi klinis yang ketat. Explainability sebaiknya diposisikan sebagai alat pendukung tata kelola dan keselamatan, bukan sebagai justifikasi tunggal untuk adopsi teknologi.

Dari sisi regulasi dan tata kelola, pendekatan lifecycle regulation menjadi semakin relevan seiring dengan sifat adaptif dan dinamis dari sistem AI. Pedoman pelaporan seperti CONSORT-AI, SPIRIT-AI, TRIPOD, dan TRIPOD+AI merupakan fondasi penting untuk meningkatkan kualitas bukti ilmiah dan memastikan bahwa sistem AI diagnostik dikembangkan serta digunakan secara bertanggung jawab.

Secara keseluruhan, kajian ini menyimpulkan bahwa masa depan diagnosis berbasis AI terletak pada kolaborasi manusia–AI, bukan pada substitusi klinisi. AI paling efektif ketika digunakan sebagai clinical decision support system yang meningkatkan akurasi, konsistensi, dan efisiensi diagnosis, sambil tetap mempertahankan peran sentral tenaga medis dalam penilaian klinis holistik, pengambilan keputusan akhir, dan tanggung jawab etis terhadap pasien.

5 Deklarasi/Pernyataan

5.1. Ucapan Terima Kasih

Para penulis menyampaikan penghargaan dan terima kasih yang sebesar-besarnya kepada Atma Jaya Teaching & Research Hospital dan Fakultas Farmasi, Universitas Mulawarman atas bimbingan akademik, dukungan institusional, serta penyediaan sumber daya yang diperlukan dalam pelaksanaan penelitian ini. Kolaborasi yang terjalin antara institusi-institusi tersebut berperan penting dalam penyelesaian karya ilmiah ini.

5.2 Konflik kepentingan

Para penulis menyatakan bahwa tidak terdapat konflik kepentingan terkait dengan publikasi artikel ini. Tidak ada kepentingan finansial maupun non-finansial, hubungan pribadi, atau afiliasi tertentu yang memengaruhi isi, analisis, maupun kesimpulan yang disajikan dalam penelitian ini. Seluruh

sumber pendanaan, apabila ada, telah diungkapkan secara transparan, dan penelitian ini dilaksanakan secara independen tanpa adanya bias komersial maupun institusional.

6 Daftar Pustaka

- [1] H. Singh, T. D. Giardina, E. N. Meyer, M. Forjuoh, R. D. Reis, and E. J. Thomas, “Types and origins of diagnostic errors in primary care settings,” *JAMA Internal Medicine*, vol. 173, no. 6, pp. 418–425, 2013.
- [2] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019.
- [3] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015.
- [4] G. Campanella et al., “Clinical-grade computational pathology using weakly supervised deep learning on whole-slide images,” *Nature Medicine*, vol. 25, no. 8, pp. 1301–1309, 2019.
- [5] A. Rajkomar et al., “Scalable and accurate deep learning for electronic health records,” *npj Digital Medicine*, vol. 1, no. 18, 2018.
- [6] G. Litjens et al., “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [7] M. Nagendran et al., “Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims,” *BMJ*, vol. 368, m689, 2020.
- [8] R. Aggarwal et al., “Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis,” *npj Digital Medicine*, vol. 4, no. 65, 2021.
- [9] M. D. Abràmoff et al., “Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices,” *npj Digital Medicine*, vol. 1, no. 39, 2018.
- [10] V. Gulshan et al., “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [11] D. Ardila et al., “End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography,” *Nature Medicine*, vol. 25, no. 6, pp. 954–961, 2019.
- [12] D. A. Bluemke et al., “Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers,” *Radiology*, vol. 294, no. 3, pp. 487–489, 2020.
- [13] S. M. McKinney et al., “International evaluation of an AI system for breast cancer screening,” *Nature*, vol. 577, pp. 89–94, 2020.
- [14] A. Y. Hannun et al., “Cardiologist-level arrhythmia detection with convolutional neural networks,” *Nature Medicine*, vol. 25, pp. 65–69, 2019.
- [15] Z. Obermeyer and E. J. Emanuel, “Predicting the future — Big data, machine learning, and clinical medicine,” *New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016.
- [16] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, “Dissecting racial bias in an algorithm used to manage the health of populations,” *Science*, vol. 366, no. 6464, pp. 447–453, 2019.
- [17] T. J. Hwang, A. S. Kesselheim, and K. N. Vokinger, “Lifecycle regulation of artificial intelligence-based medical devices,” *JAMA*, vol. 322, no. 23, pp. 2285–2286, 2019.
- [18] X. Liu et al., “CONSORT-AI: Extension of the CONSORT statement for artificial intelligence trials,” *BMJ*, vol. 370, m3164, 2020.
- [19] S. Cruz Rivera et al., “SPIRIT-AI extension for clinical trial protocols,” *Nature Medicine*, vol. 26, pp. 1351–1363, 2020.
- [20] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, “TRIPOD: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis,” *Annals of Internal Medicine*, vol. 162, no. 1, pp. 55–63, 2015.
- [21] G. S. Collins et al., “TRIPOD+AI: A reporting guideline for clinical prediction models developed using artificial intelligence,” *BMJ*, vol. 385, e078378, 2024.
- [22] E. W. Steyerberg et al., “Assessing the performance of prediction models: calibration,” *Statistics in Medicine*, vol. 29, no. 27, pp. 2869–2884, 2010.

- [23] R. F. Wolff et al., “PROBAST: A tool to assess the risk of bias and applicability of prediction model studies,” *Annals of Internal Medicine*, vol. 170, no. 1, pp. 51–58, 2019.
- [24] C. Patrício et al., “Explainable deep learning models in medical imaging: A survey,” *ACM Computing Surveys*, 2023.
- [25] B. N. Green, C. D. Johnson, and A. Adams, “Writing narrative literature reviews for peer-reviewed journals: secrets of the trade,” *Journal of Chiropractic Medicine*, vol. 5, no. 3, pp. 101–117, 2006.