

Validasi Klinis dan Penerapan Model Berbasis Kecerdasan Buatan untuk Deteksi dan Prediksi Reaksi Merugikan Obat: Tinjauan Pustaka

Clinical Validation and Application of an Artificial Intelligence-Based Model for Detection and Prediction of Adverse Drug Reactions: A Literature Review

Aditya Manggala Putra^{1*}, Alvin Dwi Reza², Rangga Febriyana³

^{1,2,3} Department of Military Pharmacy, Faculty of Military Pharmacy, The Republic of Indonesia Defense University, Sentul, 16810, Indonesia, manggalapaditya@gmail.com

Abstrak

Latar Belakang: Efek samping obat (adverse drug events, ADE) berkontribusi terhadap hampir 50% kerugian medis yang dapat dicegah secara global dan menjadi beban bagi sistem kesehatan. Farmakovigilans konvensional berbasis pelaporan spontan menghadapi underreporting, kualitas data rendah, dan keterlambatan deteksi sinyal keamanan. Kecerdasan buatan (artificial intelligence/AI) dan machine learning (ML) semakin diterapkan untuk mengatasi keterbatasan tersebut, tetapi bukti efektivitas dan hambatan implementasinya masih tersebar. **Tujuan:** Menganalisis kemajuan, efektivitas, manfaat, dan tantangan penerapan AI dalam farmakovigilans, meliputi deteksi sinyal, prediksi ADE, ekstraksi informasi klinis, evaluasi kausalitas, dan integrasi regulasi. **Metodologi:** Penelitian menggunakan tinjauan naratif terhadap artikel mengenai AI dan ML dalam farmakovigilans dan prediksi toksisitas obat. Literatur mencakup tinjauan sistematis, scoping review, studi model prediktif, dan dokumen kebijakan WHO. **Hasil:** Metode statistik seperti PRR, ROR, BCPNN, dan MGPS masih digunakan karena sederhana dan dapat diaudit, meskipun sensitivitas terbatas. Model ML seperti Random Forest, XGBoost, regresi logistik, serta transformer seperti BERT dan GPT menunjukkan kinerja menjanjikan dalam deteksi dan prediksi ADE, dengan akurasi hingga 94%. Namun, validasi eksternal, ketidakseimbangan kelas, dan interpretabilitas masih menjadi tantangan. **Kesimpulan:** AI berpotensi meningkatkan efisiensi dan akurasi farmakovigilans, tetapi memerlukan peningkatan kualitas data, validasi dunia nyata, transparansi model, standardisasi metodologi, dan integrasi dalam alur kerja klinis secara berkelanjutan.

Kata kunci : Kecerdasan Buatan, Deteksi Reaksi Obat yang Merugikan, Prediksi, Validasi Klinis

Abstract

Background : Adverse drug events (ADEs) account for nearly 50% of preventable medical harm globally and place a burden on health care systems. Conventional pharmacovigilance based on spontaneous reporting faces challenges such as underreporting, low data quality, and delays in detecting safety signals. Artificial intelligence (AI) and machine learning (ML) are increasingly being applied to address these limitations, but evidence of their effectiveness and barriers to implementation remain scattered. **Objective :** To analyze the progress, effectiveness, benefits, and challenges of applying AI in pharmacovigilance, including signal detection, ADE prediction, clinical information extraction, causality assessment, and regulatory integration. **Methodology :** This study employed a narrative review of articles on AI and ML in pharmacovigilance and drug toxicity prediction. The literature included systematic reviews, scoping reviews, predictive modeling studies, and WHO policy documents. **Results:** Statistical methods such as PRR, ROR, BCPNN, and MGPS are still used because they are simple and auditable, despite their limited sensitivity. ML models such as Random Forest, XGBoost, logistic regression, and transformers like BERT and GPT show promising performance in the detection and prediction of ADEs, with accuracies of up to 94%. However, external validation, class imbalance, and interpretability remain challenges. **Conclusion :** AI has the potential to improve the efficiency and accuracy of pharmacovigilance, but requires improved data quality, real-world validation, model transparency, methodological standardization, and ongoing integration into clinical workflows.

Keywords: Artificial Intelligence, Adverse Drug Reaction Detection, Prediction, Clinical Validation

*Corresponding author: Aditya Manggala Putra, Department of Military Pharmacy, Faculty of Military Pharmacy, The Republic of Indonesia Defense University, Sentul, 16810, Indonesia.

E-mail : manggalapaditya@gmail.com

Doi : 10.35451/jfm.v9i1.3944

Received : September 22, 2026, Accepted: October 04, 2026, Published: October 30, 2026

Copyright: © 2026 Aditya Manggala Putra (s). Creative Commons License This work is licensed under a Creative Commons Attribution 4.0 International License.

1. PENDAHULUAN

Efek samping obat (*ADE*) adalah cedera yang terjadi akibat tindakan medis terkait obat, meliputi reaksi obat yang tidak salah (*ADR*) serta komplikasi akibat kesalahan dalam pengobatan, dan berkontribusi hampir 50% dari semua kerugian medis yang bisa dicegah secara global, dengan prevalensi mencapai satu dari setiap 20 pasien [1]. *ADE* juga mengakibatkan 700.000 kunjungan darurat setiap tahun di Amerika Serikat dengan biaya sebesar 30,1 miliar dolar, serta berkontribusi pada 10.000–22.000 kematian setiap tahun di Kanada [2]. Farmakovigilans memiliki peran yang krusial dalam mengidentifikasi, mengevaluasi, dan mencegah *ADR*, tetapi metode konvensional yang tergantung pada laporan kasus keamanan individu rentan terhadap bias serta masalah underreporting yang signifikan, sehingga dianggap semakin tidak memadai dalam menghadapi kompleksitas data keamanan obat saat ini [3-4].

Kecerdasan buatan (*AI*) dan pembelajaran mesin (*ML*) saat ini bertransformasi dari aplikasi percobaan menjadi penggunaan sehari-hari dalam farmakovigilans, sejak diluncurkannya algoritma deteksi sinyal seperti BCPNN dan MGPS di awal tahun 2000-an [4]. Ketersediaan data nyata mendorong penggunaan *AI* untuk mengenali pola rumit yang sulit dijangkau oleh metode statistik tradisional [3]. Walaupun demikian, masih ada kekurangan kejelasan terkait pelaksanaan dan validasi model *ML* dalam prediksi *ADE*, dengan keragaman metodologis yang ada serta interpretabilitas dan validasi eksternal yang sering dianggap tidak memadai, sehingga menyulitkan pengambilan kesimpulan sistematis mengenai praktik terbaik [2,5]. Tenaga kesehatan yang berfungsi sebagai pelapor utama memiliki peran penting, karena pengetahuan, sikap, dan praktik mereka berpengaruh pada mutu pelaporan *ADR*, sedangkan *underreporting* masih terjadi secara global dengan hanya 6–10% *ADR* yang telah dilaporkan [6]. Dengan demikian, penelitian ini bertujuan untuk mengkaji kemajuan, efektivitas, keunggulan, dan tantangan implementasi kecerdasan buatan dalam mendukung sistem farmakovigilans, memberikan pandangan menyeluruh bagi peneliti, klinisi, dan pengambil kebijakan dalam memperkuat keamanan pemakaian obat di masa depan.

2. METODE

Strategi Pencarian

Penelitian ini mengadopsi pendekatan tinjauan literatur dengan melakukan pencarian referensi pada basis data PubMed dan Google Scholar. Pencarian dilakukan secara terencana dengan memanfaatkan kombinasi kata kunci serta operator *Boolean AND* dan *OR*. Strategi pencarian yang diterapkan mencakup: (“*Artificial Intelligence*” *OR* “*Machine Learning*” *OR* “*Deep Learning*” *OR* “*Artificial Neural Network*”) *AND* (“*Adverse Drug Reaction*” *OR* “*Adverse Drug Event*” *OR* “*Adverse Reaction*”) *AND* (“*Detection*” *OR* “*Prediction*” *OR* “*Identification*”) *AND* (“*Clinical Validation*” *OR* “*External Validation*” *OR* “*Clinical Applicability*” *OR* “*Clinical Utility*”). Literatur yang diperhatikan adalah artikel riset yang diterbitkan dalam periode 5 tahun terakhir, yaitu dari tahun 2021 hingga Agustus 2026.

Teknik Pengumpulan Data

Artikel yang diperoleh dari basis data PubMed dan Google Scholar dikumpulkan, selanjutnya ditinjau dan diidentifikasi sesuai dengan tujuan penelitian. Studi dinilai memenuhi syarat jika membahas penggunaan *artificial intelligence (AI)* atau *machine learning (ML)* untuk mendeteksi dan/atau memprediksi *adverse drug reaction (ADR)* serta mencakup hasil mengenai kinerja model, validasi, atau penerapannya dalam konteks klinis. Kriteria inklusi meliputi studi asli pada manusia yang membahas pemanfaatan *AI* atau *ML* untuk mendeteksi atau memprediksi *ADR*, memiliki hasil evaluasi kinerja model atau validasi, serta diterbitkan dalam periode tahun 2021–2026. Kriteria eksklusi mencakup artikel non-penelitian asli seperti editorial, surat, prosiding, dan opini, artikel yang tidak membahas *ADR* atau *AI* atau *ML* sesuai dengan tujuan penelitian, studi pada hewan atau *in vitro*, serta artikel yang diterbitkan di luar kisaran tahun 2021–2026. Artikel yang memenuhi kriteria selanjutnya dianalisis secara mendetail untuk mengumpulkan informasi tentang karakteristik studi, jenis algoritme *AI*, sumber data, populasi penelitian, metode validasi, kinerja prediktif, dan aplikasi klinis.

Ekstraksi Data

Ekstraksi dilakukan setelah judul, abstrak beserta isi teks disaring, penulis melakukan penilaian secara independen dan menganalisis studi yang dipilih. Data yang diambil dari studi setelah diidentifikasi diantaranya, Model *AI* yang dibahas pada studi, Efektivitas dan akurasi terdapat pada studi, Populasi yang menjadi sampel pada studi, Hasil kelebihan dan tantangan studi, Tahun studi tersebut dilaksanakan

3. HASIL

Studi ini mengambil studi melalui Pubmed dan Google Scholar dengan jumlah data studi yang belum melalui fase ekstraksi sebanyak 250 artikel. Artikel sebanyak 250 tersebut melalui proses ekstraksi yang sudah ditentukan kriteria hasil yang akan diambil hingga menghasilkan jumlah akhir sebanyak 11 artikel yang akan diidentifikasi dan ditinjau lebih lanjut.

No	Judul	Tahun/Penulis	Populasi/Sumber Data	Model AI yang Digunakan	Efektivitas & Akurasi	Kelebihan	Tantangan
1	Kajian Kejadian Efek Samping Obat (ESO) pada Pasien Geriatri di Ruang Rawat Inap Interne Rumah Sakit Swasta Bukittinggi [7]	2024 / Jannah, Sari, Almasdy	12 informan (dokter, perawat, apoteker) di 1 RS swasta Bukittinggi, Sumbar; wawancara semi-terstruktur, dianalisis dengan NVivo 12	Bukan model AI prediktif — analisis kualitatif content analysis berbantuan software (NVivo)	N/A (studi kualitatif, tidak melaporkan akurasi numerik). Temuan: obat H2 blocker & antibiotik paling sering menyebabkan ESO kulit	Menyelidiki persepsi dan pengalaman tenaga kesehatan secara mendalam; landasan kontekstual untuk pengembangan sistem farmakovigilans digital lokal.	Sampel terbatas, satu rumah sakit, tidak menerapkan AI kuantitatif; pemahaman farmakovigilans nakes masih minim, pelaporan ESO belum optimal.
2	Digital Monitoring of Medication Safety in Children: ADR Signalling Techniques in Malaysia [8]	2024 / Rekha, Hisham, Wahab, dkk.	3152 laporan ADR anak (0-12 th), database NPRA Malaysia 2016-2020	4 teknik disproporsionalitas: PRR, ROR, BCPNN, MGPS (algoritma statistik farmakovigilans)	MGPS: sensitivitas tertinggi 20%, spesifisitas 100%, PPV 100%, NPV 23%; PRR/ROR/BCPNN: sensitivitas 12%, spesifisitas 100%	Spesifisitas & PPV tinggi di semua metode -> sinyal yang terdeteksi sangat mungkin sah; MGPS mendeteksi kelima sinyal keamanan.	Sensitivitas rendah (kumpulan data kecil, pelaporan yang kurang); tidak dapat membuktikan kausalitas secara langsung; memerlukan database >500 laporan untuk hasil yang stabil.
3	Develop an ADR Prediction System of Chinese Herbal Injections Containing Panax Notoginseng Saponin [9]	2022 / Wu, Zhang, Chang, dkk. (BMJ Open)	530 pasien (5 RS Sichuan, 2010-2018), studi nested case-control	18 algoritma ML: AdaBoost, Bagging, Naive Bayes, Decision Tree, RF, Gradient Boosting, KNN, SVM, XGBoost, Ensemble Learning, dll (1080 model)	Model terbaik: AUC 0.9141, akurasi 89.47%, presisi 0.75, recall 0.6667, F1 0.7059	Sistem prediksi ADR secara real-time yang telah memiliki hak cipta perangkat lunak; SHAP untuk penjelasan variabel signifikan (fungsi ginjal, dermatosis, usia).	Sampel kecil dan hanya berasal dari satu wilayah (Sichuan) -> generalisasi terbatas; memerlukan uji coba klinis prospektif.

4	Artificial Intelligence in Pharmacovigilance: Narrative Review & Bayesian Network Tool [10]	2025 / Algarvio, Conceicao, Rodrigues, Ribeiro, Ferreira-da-Silva	Literatur MEDLINE, Scopus, Web of Science; studi kasus di Porto Pharmacovigilance Centre (data 12 tahun)	Bayesian Network (expert-defined) untuk causality assessment, dibandingkan dengan ML, DL, NLP secara umum	PPV level Definite 71.4%, level Probable 87.3%; waktu proses turun dari hari menjadi jam	Mengurangi bias dalam evaluasi kausalitas; sejalan dengan pendapat para pakar; mempercepat proses tanpa mengesampingkan keahlian manusia.	Mewarisi bias dari standar acuan (gold standard); tidak terbaru secara otomatis dengan kasus-kasus baru; antarmuka tidak cukup ramah pengguna.
5	Artificial Intelligence-Powered Pharmacovigilance: Review of ML/DL dalam Deteksi ADE Berbasis Teks Klinis [11]	2024 / Li, Tao, Li, Sun, dkk. (J Biomed Inform)	Dataset benchmark: 2018 n2c2 Shared Task & 2018 MADE 1.0 Challenge (catatan klinis EHR)	CRF, BiLSTM-CRF, CNN-BiLSTM-CRF, BERT, BioBERT, SciBERT, RoBERTa, BlueBERT, SVM, RF, Gradient Boosting	Tugas NER: F1 terbaik 0.94 (SciBERT/BlueBERT); Tugas RC: F1 0.96 (Gradient Boosting); End-to-end terbaik (BERT): F1 0.94	Perbandingan sistematis berbagai model pada dataset standar; keluarga BERT lebih unggul dalam menangkap hubungan jarak jauh dalam teks klinis.	Entitas 'ADE' dan 'Reason' sulit untuk diekstraksi (data tidak seimbang, hanya 2-8% entitas); model memerlukan data anotasi yang besar dan mahal dalam komputasi.
6	Key Use Cases for AI to Reduce the Frequency of Adverse Drug Events: A Scoping Review [12]	2022 / Syrowatka, Song, Amato, dkk. (Lancet Digital Health)	78 artikel (1998-2020), populasi dari Amerika Utara, Eropa, Asia; ukuran sampel 200 - >1.000.000 pasien	Neural network, Random Forest, SVM, Decision Tree, Bayesian, KNN, Gradient Boosting Machine (ML/DL modern)	Sebagian besar studi hanya melaporkan performa teknis (akurasi/AUC-ROC bervariasi luas); hanya 5 studi (6%) menguji efikasi klinis nyata	Mengidentifikasi 4 kasus penggunaan prediksi (risiko ADE, respons terapi, dosis optimal, pilihan terapi) & 2 kasus penggunaan deteksi awal.	Sebagian besar penelitian masih berada di tahap awal (fase 0-1 uji klinis); metrik akurasi/AUC-ROC tidak cukup akurat untuk ADE yang jarang; integrasi data genetik dan catatan yang tidak terstruktur masih terbatas.

7	AI-Driven Pharmacovigilance: Enhancing ADR Detection with Deep Learning and NLP [13]	2025 / Khemani, Malave, Shinde, dkk. (MethodsX)	Data uji klinis terstruktur & tak terstruktur (demografi, lab, catatan klinis, EHR, data SMILES obat baru)	Logistic Regression, RF, XGBoost, CNN, BERT, GPT, k-Means, DBSCAN, kombinasi CNN+BERT+k-Means	GPT: akurasi 94.1% (tertinggi); CNN+BERT+k-Means: 91%; BERT: 92.7%; CNN: 89%; XGBoost: 88.2%; Logistic Regression: 78.5%	Model multi-modal hibrida (struktur + teks); dilengkapi dengan SHAP & LIME untuk interpretasi; validasi menggunakan 5-fold cross-validation & pencegahan overfitting.	GPT paling tepat tetapi lama dalam waktu inferensi (tidak ideal untuk real-time); memerlukan komputasi besar; hanya terverifikasi pada data uji klinis yang terstruktur.
8	Computational Toxicology in Drug Discovery: AI dalam ADMET & Toxicity Prediction [14]	2025 / Zhang, Li, Zhang, dkk. (Briefings in Bioinformatics)	>20 platform prediksi ADMET & 4 jenis database toksikologi (kimia, lingkungan, alternatif, toksin biologis) - review literatur	Metode berbasis aturan/statistik, ML (RF, SVM, dll.), graph-based (GNN), model generatif, LLM	Tidak melaporkan satu angka akurasi tunggal (review); membahas tren pergeseran dari prediksi single-endpoint ke multi-endpoint/multimodal	Ruang lingkup yang luas (akut, spesifik organ, karsinogenitas); membahas kerangka kerja interpretabilitas & LLM untuk eksplorasi literatur toksikologi.	Kualitas data, interpretasi model, dan inferensi kausal tetap menjadi tantangan utama; belum terdapat standar evaluasi konsisten di berbagai platform.
9	Review of Machine Learning and Deep Learning Models for Toxicity Prediction [15]	2023 / Guo, Liu, Dong, dkk. (Experimental Biology and Medicine, FDA/NCTR)	Berbagai dataset toksisitas kimia/obat (hepatotoksitas, kardiotoxsisitas, karsinogenisitas) - review literatur	SVM, Random Forest, KNN, Ensemble Learning, Deep Neural Network	SVM dan RF paling sering digunakan & berkinerja baik; performa sangat bergantung pada kualitas dataset (bervariasi antar dataset utk toksisitas sama)	Menganalisis endpoint toksisitas yang paling sering dimodelkan; menekankan pentingnya dataset benchmark yang konsisten.	Variasi label toksisitas antara dataset untuk senyawa yang serupa; ketidakadaan dataset standar dengan kualitas tinggi mengurangi keandalan model.
10	Machine Learning-Enabled Drug-Induced Toxicity Prediction [16]	2025 / Bai, Wu, Li, Cao, He, Bo	10 kategori toksisitas akibat obat; berbagai database toksikologi,	Model prediktif & model interpretable ML (perbandingan	Tidak melaporkan akurasi tunggal; menekankan bahwa	Analisis perbandingan luas 10 tipe toksisitas	Beragamnya sumber data membuat peneliti

		(Advanced Science)	kimia, multi-omics, & benchmark - review literatur	lintas kategori toksisitas, tidak dirinci satu model tunggal)	model optimal berbeda-beda tergantung jenis toksisitas	obat; menghubungkan ramalan dengan pemahaman mekanistik.	kesulitan dalam memusatkan perhatian pada toksisitas tertentu; belum ada satu model AI yang dapat diandalkan untuk semua tipe toksisitas.
11	Robustness in Deep Learning Models for Medical Diagnostics: Security and Adversarial Challenges [17]	2024 / Javed, El-Sappagh, Abuhmed (Artificial Intelligence Review)	Tinjauan literatur luas model diagnostik medis berbasis DL (berbagai penyakit: infeksi, genetik, autoimun, dll.)	Deep Learning secara umum; dibahas teknik pertahanan: adversarial training, input preprocessing, data augmentation, uncertainty estimation	Tidak melaporkan akurasi spesifik (bukan studi prediksi tunggal); fokus pada ketahanan model terhadap input adversarial/noise	Membahas metrik penilaian ketahanan dan alat (TensorFlow, PyTorch) yang berkaitan untuk keandalan sistem AI medis.	Model yang kompleks lebih mudah mengalami overfitting dan serangan adversarial; kualitas dan kuantitas data pelatihan memiliki dampak besar pada keandalan; ada risiko keamanan dan privasi (pengambilan data sensitif).

4. PEMBAHASAN

Perubahan Statistik dari Statistik Klasik menjadi *Deep Learning*

Tabel hasil menunjukkan adanya spektrum metode yang terbagi dengan jelas menjadi tiga generasi. Generasi pertama mencakup teknik disproporsionalitas statistik klasik seperti PRR, ROR, BCPNN, dan MGPS yang diterapkan oleh mengenai informasi anak-anak Malaysia [8]. Metode ini tetap menjadi acuan bagi lembaga pengatur (WHO UMC, FDA, BPOM) karena mudah, bisa diaudit, dan tidak memerlukan data pelatihan yang besar. Namun, hasil dalam tabel mengungkapkan keterbatasan dasarnya: sensitivitas yang sangat rendah (12 hingga 20%) meskipun spesifisitas dan PPV mencapai 100%. Dengan kata lain, metode ini jarang menghasilkan sinyal palsu, tetapi juga sering kali melewatkan sinyal *ADR* yang benar-benar ada, sebagai akibat langsung dari underreporting dan ukuran basis data yang terbatas.

Generasi kedua mencakup machine learning klasik (Random Forest, SVM, XGBoost, Gradient Boosting, Naïve Bayes, dan lain-lain), yang merupakan fokus utama dalam penelitian dan berfungsi sebagai referensi di sebagian besar artikel lainnya [9]. Generasi ketiga mencakup deep learning berbasis transformer (BERT, BioBERT, SciBERT, BlueBERT, GPT), yang secara konsisten menunjukkan kinerja paling baik dalam mengekstrak informasi dari teks klinis yang tidak terstruktur, seperti yang ditunjukkan dalam penelitian, dengan skor F1 atau tingkat akurasi antara 0,94 dan 0,941 [11,13].

Pola ini menegaskan hasil yang diperoleh oleh bahwa *AI* bukan menggantikan metode tradisional, tetapi sebaliknya melengkapi [10]. Statistik disproporsionalitas tepat untuk triase yang cepat dan transparan pada basis data kecil hingga menengah, sementara deep learning lebih unggul saat data tersedia dalam jumlah besar dan dalam bentuk teks bebas (catatan klinis, laporan naratif *ADR*).

Keseimbangan antara Presisi dan Kelayakan Praktis

Salah satu hasil signifikan yang selalu muncul dalam berbagai artikel adalah adanya trade-off antara ketepatan model dan kemudahan penerapannya di lapangan [12-13].

GPT mencatat tingkat akurasi tertinggi (94,1%) tetapi juga waktu inferensi terlama, sehingga kurang sesuai untuk deteksi *ADR* secara real-time di rumah sakit. CNN dan XGBoost memberikan kombinasi yang lebih baik antara akurasi (88 hingga 89%) dan kecepatan pemrosesan, sehingga keduanya menjadi pilihan yang lebih praktis untuk sistem pendukung keputusan klinis (CDSS) yang perlu memberikan respons cepat.

Model klasik seperti Regresi Logistik masih yang tercepat, namun akurasinya paling rendah (sekitar 78%). Ini menunjukkan bahwa "model terbaik" tidak hanya model dengan akurasi tertinggi, tetapi model yang paling cocok dengan konteks pemakaian, apakah untuk pengenalan sinyal retrospektif (di mana akurasi lebih diutamakan) atau untuk peringatan klinis waktu nyata (di mana kecepatan dan interpretasi lebih penting).

Perbedaan Nyata antara Performa Teknis dan Validasi Nyata

Temuan yang paling penting berasal dari tinjauan ruang lingkup dari 78 penelitian *AI* terkait pengurangan *ADE*, 94% hanya menguji kinerja teknis algoritma, dan hanya 5 studi (6%) yang menilai efikasi *AI* dalam konteks klinis nyata (setara fase 2 uji klinis) [12]. Pola yang sama terlihat di hampir semua artikel lain dalam tabel, termasuk, yang semuanya masih berada pada fase pengembangan dan validasi model (fase 0 hingga 1), bukan pada fase implementasi dan pengujian dampak klinis [9,11,13].

Ini mengindikasikan terdapatnya jurang riset yang penting. Kemampuan teknis *AI* dalam mendeteksi atau memprediksi *ADR* telah cukup berkembang, namun data yang menunjukkan bahwa *AI* benar-benar mempengaruhi perilaku persepsian dokter, menurunkan kejadian *ADR* di lapangan, atau efisien secara biaya masih sangat sedikit. Terdapat pengecualian yang menarik karena telah secara teratur menerapkan Jaringan Bayes di Pusat Farmakovigilans Porto selama bertahun-tahun, dan berhasil menunjukkan penurunan durasi proses penilaian kausalitas dari hari menjadi jam, menjadikannya salah satu dari sedikit contoh nyata dampak operasional *AI* dalam farmakovigilans [10].

Tantangan Penerapan

Walaupun ada variasi dalam metode, hampir semua artikel dalam tabel mencatat kendala yang serupa: Ketidakseimbangan kelas (class imbalance): entitas semacam "*ADE*" atau "*Reason*" hanya terdiri dari 2 hingga 8% dari keseluruhan data anotasi, sehingga model mengalami kesulitan dalam mempelajari pola umum untuk situasi yang jarang terjadi [11].

Underreporting: baik penelitian. secara jelas menunjukkan bahwa pelaporan *ADR* di lembaga kesehatan masih kurang ideal, mengakibatkan basis data yang dipakai untuk melatih model *AI* sudah mengalami bias sejak awal [8,7].

Keterbatasan generalisasi: model yang dilatih di satu area atau institusi tertentu berpotensi tidak menunjukkan kinerja yang sama saat diterapkan pada populasi atau sistem kesehatan lainnya, seperti yang juga dinyatakan [9,12].

Keterbatasan interpretabilitas (karakteristik kotak hitam): ditelaah secara mendalam oleh Javed dan rekan-rekan serta menjadi fokus utama dalam penelitian Khemani dan rekan-rekan, yang kemudian menggabungkan SHAP dan LIME untuk memberikan penjelasan tentang keputusan model.

Pemahaman Penjelasan Proses sebagai Syarat Penerimaan Klinis

Menariknya, artikel-artikel yang paling mengutamakan interpretabilitas dengan Jaringan Bayesian yang jelas adalah yang paling relevan dengan penerapan nyata [10,13]. Hal ini menunjukkan bahwa dalam konteks kesehatan, di mana keputusan terkait keselamatan pasien, hanya memiliki akurasi tinggi tidaklah memadai. Tenaga kesehatan membutuhkan model yang dapat dipahami, selaras dengan penemuan Javed dan rekan-rekan bahwa sifat kotak hitam dari model deep learning masih menjadi kendala kepercayaan klinis, meskipun akurasinya sangat tinggi.

Perluasan Lingkup dari *ADR* ke Toksisitas Praklinis

Tiga artikel terakhir memperluas pandangan dengan menunjukkan bahwa tantangan serupa, seperti kualitas data, minimnya standar acuan, serta tidak adanya model tunggal yang superior di semua jenis toksisitas, juga terdapat dalam bidang prediksi toksisitas praklinis (ADMET) [16,15,14]. Hal ini menegaskan bahwa isu utama implementasi *AI* dalam keamanan obat bersifat melintasi berbagai tahap. Mulai dari penelitian praklinis, penelitian klinis, hingga pemantauan setelah pemasaran (farmakovigilans), tantangan utama tetap serupa, yaitu akses terhadap data bermutu dan validasi yang dapat diandalkan.

5. KESIMPULAN

Hasil tinjauan menunjukkan bahwa *AI*, terutama model yang menggunakan transformer seperti BERT dan GPT, memiliki potensi signifikan dalam meningkatkan sensitivitas serta efisiensi deteksi *ADR* dibandingkan metode statistik tradisional. Namun, potensi ini masih terhalang oleh tiga kesenjangan utama: (1) kurangnya validasi di lingkungan klinis nyata, (2) masalah kualitas dan ketidakseimbangan data yang disebabkan oleh underreporting, dan (3) kebutuhan akan interpretabilitas yang belum sepenuhnya dipenuhi oleh model dengan kinerja tinggi. Oleh karena itu, pengembangan ke depan seharusnya tidak hanya menitikberatkan pada peningkatan akurasi model, tetapi juga pada integrasi sistem *AI* ke dalam alur kerja klinis yang nyata, standarisasi basis data farmakovigilans antar institusi, serta pengembangan lebih lanjut kerangka kerja explainable *AI* (XAI) agar hasil prediksi dapat diandalkan dan digunakan secara bertanggung jawab oleh para tenaga kesehatan.

UCAPAN TERIMA KASIH

Penelitian ini telah dibahas dan didukung oleh Fakultas Farmasi Militer, Universitas Pertahanan Indonesia, Bogor, Indonesia. Kami mengucapkan terima kasih kepada Fakultas Farmasi Militer, Universitas Pertahanan Indonesia, Bogor, Indonesia

DAFTAR PUSTAKA

1. World Health Organization. Medication without harm: policy brief. Geneva: World Health Organization; 2023.
2. Chalabianloo N, Sedig K, Rostamzadeh N, Muanda FT, Izzedin L. Machine learning methods for predicting adverse drug events : A systematic review. 2026;(February 2025):422–44.
3. Dimitsaki S, Natsiavas P, Jaulent M christine. Applying AI to Structured Real-World Data for Pharmacovigilance Purposes : Scoping Review Corresponding Author : 2024;26.
4. Nagar A. Artificial intelligence in pharmacovigilance : advancing drug safety monitoring and regulatory

integration. 2025;1–16.

5. Karampatea A, Kassandros K. Systematic review of AI-based models in pharmacoepidemiology for adverse drug event prediction and detection. 2026;(March):1–9.
6. Manek MS, Wiedyaningsih C, Endarti D. A Systematic Review of Healthcare Professionals ' Knowledge , Attitudes , and Practices Regarding Adverse Drug Reaction Reporting in Southeast Asia. 2026;1133–45.
7. Jannah M, Sari YO, Almasdy D. GERIATRI DI RUANGAN RAWAT INAP INTERNE SEBUAH SUMATERA BARAT Study Of the Adverse Drug Reaction (ADR) On Geriatric Patients in Internal Medicine Inpatient at A Private Hospital in Bukittinggi City , West Sumatera. 2024;6(2).
8. Rekha BH, Hisham SA, Wahab IA, Ali NM, Goh KW. Digital monitoring of medication safety in children : an investigation of ADR signalling techniques in Malaysia. 2024;
9. Wu XW, Song W, Wen L. Develop an ADR prediction system of Chinese herbal injections containing Panax notoginseng saponin : a nested case – control study using machine learning. 2022;1–8.
10. Algarvio RC, Conceição J, Pereira P, Inês R. Artificial intelligence in pharmacovigilance : a narrative review and practical experience with an expert - defined Bayesian network tool. *Int J Clin Pharm* [Internet]. 2025;47(4):932–44. Available from: <https://doi.org/10.1007/s11096-025-01975-3>
11. Li Y, Tao W, Li Z, Sun Z, Li F, Fenton S, et al. Artificial intelligence-powered pharmacovigilance : A review of machine and deep learning in clinical text-based adverse drug event detection for benchmark datasets. *J Biomed Inform* [Internet]. 2024;152(February):104621. Available from: <https://doi.org/10.1016/j.jbi.2024.104621>
12. Syrowatka A, Song W, Amato MG, Foer D, Edrees H, Co Z, et al. Review Key use cases for artificial intelligence to reduce the frequency of adverse drug events : a scoping review. 2022;
13. Khemani B, Malave S, Shinde S, Shukla M. MethodsX AI-driven pharmacovigilance : Enhancing adverse drug reaction detection with deep learning and NLP ☆. *MethodsX* [Internet]. 2025;15(June):103460. Available from: <https://doi.org/10.1016/j.mex.2025.103460>
14. Zhang J, Li H, Zhang Y, Huan J, Ren L, Zhang C, et al. Computational toxicology in drug discovery : applications of artificial intelligence in ADMET and. 2025;26(5).
15. Guo W, Liu J, Dong F, Song M, Li Z, Khan KH, et al. Minireview Review of machine learning and deep learning models for toxicity prediction. 2023;1952–73.
16. Bai C, Wu L, Li R, Cao Y, He S, Bo X. Machine Learning-Enabled Drug-Induced Toxicity Prediction. 2025;2413405.
17. El-sappagh HJS, Abuhmed T. Robustness in deep learning models for medical diagnostics : security and adversarial challenges towards robust AI. 2025;