

## **APLIKASI MACHINE LEARNING DALAM PREDIKSI INTERAKSI OBAT: SYSTEMATIC REVIEW**

**Dody May Arfian<sup>1</sup>, I Gede Irvan Pramanta Andika<sup>2</sup>, Riska Wiradarma<sup>3</sup>**

<sup>1</sup>Program Studi Informatika, Fakultas Teknologi, Institut Teknologi dan Kesehatan Bintang Persada  
 Jalan Gatot Subroto, Denpasar, Indonesia

<sup>2</sup>Program Studi Informatika, Fakultas Teknologi, Institut Teknologi dan Kesehatan Bintang Persada  
 Jalan Gatot Subroto, Denpasar, Indonesia

e-mail: [dodymayarfian@gmail.com](mailto:dodymayarfian@gmail.com)<sup>1</sup>, [irvanpramanta@gmail.com](mailto:irvanpramanta@gmail.com)<sup>2</sup>, [riskawiradarma@gmail.com](mailto:riskawiradarma@gmail.com)<sup>3</sup>

Received : January, 2026	Accepted : Maret, 2026	Published : April, 2026
--------------------------	------------------------	-------------------------

### **Abstract**

*Drug-drug interactions (DDIs) represent a significant challenge in modern pharmacotherapy, contributing to 17-23% of adverse drug reaction-related hospitalizations. Machine learning (ML) has emerged as a promising approach for computational DDI prediction, yet a comprehensive synthesis of methodologies, performance benchmarks, and clinical translation challenges remains lacking. This systematic review aims to identify and evaluate ML algorithms applied to DDI prediction, compare their performance across different datasets and validation strategies, analyze feature representation methods, and identify critical gaps impeding clinical deployment. Following PRISMA 2020 guidelines, we conducted a systematic search across five electronic databases (PubMed, IEEE Xplore, Scopus, Web of Science, Google Scholar) for studies published between January 2019 and March 2025. Dual independent screening and extraction were performed with quality assessment using adapted PROBAST criteria. Included studies were analyzed for algorithm types, feature representations, datasets, validation strategies, and performance metrics. From 1,285 initial records, 60 high-quality studies were included. Graph neural networks (GNNs) emerged as state-of-the-art methods (mean F1-score:  $0.931 \pm 0.024$ ), significantly outperforming traditional ML ( $0.842 \pm 0.038$ ,  $p < 0.001$ ) and deep neural networks ( $0.893 \pm 0.031$ ,  $p = 0.003$ ). Multi-modal approaches integrating chemical structure, biological targets, and phenotypic data achieved highest performance (F1: 0.945-0.982). DrugBank was the most utilized dataset (63.3% of studies), though severe class imbalance (positive:negative ratio 1:20 to 1:50) posed significant challenges. Critical gaps identified include: cold-start problem (18.3% performance degradation for unseen drugs), interpretability issues (45% black-box models), and minimal real-world validation (only 6.7% used EHR data). A severe reproducibility crisis was evident, with only 11.7% of studies fully reproducible. While ML-based DDI prediction has achieved impressive benchmark performance, substantial challenges remain for clinical translation. Priority research directions include: developing explainable AI methods for biological validation, addressing cold-start generalization through meta-learning and transfer learning, conducting multi-center real-world validation studies, establishing standardized evaluation protocols, and implementing federated learning infrastructure for privacy-preserving collaboration. Community-wide efforts toward reproducibility, standardization, and responsible deployment are essential for translating computational advances into clinically impactful systems that enhance medication safety.*

**Keywords:** drug-drug interaction, machine learning, deep learning, graph neural networks, systematic review, pharmacovigilance, clinical decision support, explainable AI

## Abstrak

Interaksi obat-obat (DDI) merupakan tantangan signifikan dalam farmakoterapi modern, berkontribusi terhadap 17-23% hospitalisasi terkait reaksi obat merugikan. *Machine learning* (ML) telah muncul sebagai pendekatan yang menjanjikan untuk prediksi DDI secara komputasional, namun sintesis komprehensif mengenai metodologi, patokan performa, dan tantangan translasi klinis masih kurang. *Systematic review* ini bertujuan untuk mengidentifikasi dan mengevaluasi algoritma ML yang diterapkan untuk prediksi DDI, membandingkan performa mereka di berbagai dataset dan strategi validasi, menganalisis metode representasi fitur, dan mengidentifikasi kesenjangan kritis yang menghambat penerapan klinis. Mengikuti pedoman PRISMA 2020, kami melakukan pencarian sistematis di lima basis data elektronik (PubMed, IEEE Xplore, Scopus, Web of Science, Google Scholar) untuk studi yang dipublikasikan antara Januari 2019 hingga Maret 2025. Penyaringan dan ekstraksi independen ganda dilakukan dengan penilaian kualitas menggunakan kriteria PROBAST yang diadaptasi. Studi yang diinklusi dianalisis untuk jenis algoritma, representasi fitur, dataset, strategi validasi, dan metrik performa. Dari 1.285 rekaman awal, 60 studi berkualitas tinggi diinklusi. *Graph neural networks* (GNN) muncul sebagai metode terkini (*F1-score* rata-rata:  $0,931 \pm 0,024$ ), secara signifikan melampaui ML tradisional ( $0,842 \pm 0,038$ ,  $p < 0,001$ ) dan *deep neural networks* ( $0,893 \pm 0,031$ ,  $p = 0,003$ ). Pendekatan *multi-modal* yang mengintegrasikan struktur kimia, target biologis, dan data fenotipik mencapai performa tertinggi (*F1*:  $0,945-0,982$ ). DrugBank merupakan dataset yang paling banyak digunakan (63,3% studi), meskipun ketidakseimbangan kelas yang parah (rasio positif:negatif 1:20 hingga 1:50) menimbulkan tantangan signifikan. Kesenjangan kritis yang teridentifikasi meliputi: masalah *cold-start* (penurunan performa 18,3% untuk obat yang belum pernah dilihat), isu interpretabilitas (45% model *black-box*), dan minimnya validasi dunia nyata (hanya 6,7% menggunakan data EHR). Krisis reproduksibilitas yang parah terlihat jelas, dengan hanya 11,7% studi yang sepenuhnya dapat direproduksi. Meskipun prediksi DDI berbasis ML telah mencapai performa acuan yang mengesankan, tantangan substansial masih ada untuk translasi klinis. Arah penelitian prioritas meliputi: pengembangan metode *explainable AI* untuk validasi biologis, penanganan generalisasi *cold-start* melalui *meta-learning* dan *transfer learning*, pelaksanaan studi validasi dunia nyata multi-pusat, penetapan protokol evaluasi terstandarisasi, dan implementasi infrastruktur *federated learning* untuk kolaborasi yang menjaga privasi. Upaya seluruh komunitas menuju reproduksibilitas, standarisasi, dan penerapan yang bertanggung jawab sangat penting untuk menerjemahkan kemajuan komputasional menjadi sistem yang berdampak klinis yang meningkatkan keamanan pengobatan.

**Kata Kunci:** interaksi obat-obat, *machine learning*, *deep learning*, *graph neural networks*, *systematic review*, farmakovigilans, sistem pendukung keputusan klinis, *explainable AI*

## 1. PENDAHULUAN

Drug-drug interaction (DDI) merupakan fenomena di mana efek suatu obat berubah akibat konsumsi obat lain secara bersamaan [6]. DDI menjadi isu penting dalam praktik klinis karena berkontribusi signifikan terhadap adverse drug reactions (ADR) yang dapat menyebabkan hospitalisasi bahkan kematian [3]. Risiko ini semakin meningkat seiring dengan praktik polifarmasi, terutama pada pasien lanjut usia dan penderita komorbiditas, di mana penggunaan banyak obat secara bersamaan meningkatkan kemungkinan terjadinya interaksi [4].

Secara mekanistik, DDI melibatkan interaksi farmakokinetik (absorpsi, distribusi, metabolisme, ekskresi) dan farmakodinamik (mekanisme aksi obat), yang dapat menghasilkan efek sinergis, antagonis, atau aditif [6]. Dampak klinisnya bervariasi, mulai dari penurunan efektivitas terapi hingga toksisitas serius, serta berkontribusi terhadap kegagalan terapi dan peningkatan angka readmisi rumah sakit [3]. Selain itu, beban ekonomi akibat DDI juga sangat besar dan menjadi perhatian dalam sistem layanan kesehatan modern [4].

Metode konvensional dalam identifikasi DDI, seperti uji *in vitro* dan *in vivo*, memiliki keterbatasan signifikan. Prosesnya

membutuhkan waktu dan biaya besar, serta tidak mampu menguji seluruh kemungkinan kombinasi obat yang jumlahnya sangat besar [4]. Selain itu, banyak interaksi baru terdeteksi pada fase post-marketing, dan metode tradisional sering kali tidak mampu menjelaskan mekanisme molekuler secara mendalam [1].

Perkembangan artificial intelligence (AI) dan machine learning (ML) menawarkan pendekatan alternatif yang lebih efisien dan skalabel dalam prediksi DDI [2]. Metode ML mampu menganalisis data biomedis dalam jumlah besar, mengidentifikasi pola kompleks, serta memprediksi interaksi obat yang belum pernah diuji secara eksperimental [1][2]. Berbagai pendekatan telah dikembangkan, mulai dari algoritma klasik seperti Support Vector Machine dan Random Forest hingga metode deep learning seperti Convolutional Neural Networks, Recurrent Neural Networks, dan Graph Neural Networks [4]. Pendekatan berbasis graf menunjukkan performa tinggi karena mampu merepresentasikan struktur molekul dan jaringan biologis secara alami [2][8].

Meskipun demikian, penerapan ML dalam prediksi DDI masih menghadapi berbagai tantangan, seperti keterbatasan data berlabel, ketidakseimbangan kelas, masalah cold-start pada obat baru, serta rendahnya interpretabilitas model [4][8]. Selain itu, masih terdapat kesenjangan antara performa model komputasional dengan validasi klinis, serta belum adanya standarisasi dalam dataset dan metodologi evaluasi [1].

Berdasarkan hal tersebut, diperlukan kajian sistematis yang komprehensif untuk mengidentifikasi perkembangan terkini, membandingkan performa berbagai metode, serta menganalisis tantangan dan peluang penelitian di masa depan. Tinjauan ini diharapkan dapat mendukung pengembangan sistem prediksi DDI yang lebih akurat, dapat diinterpretasikan, dan aplikatif dalam praktik klinis [1][2].

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan **systematic literature review (SLR)** untuk mengidentifikasi, mengevaluasi, dan mensintesis bukti ilmiah terkait penerapan machine learning dalam prediksi drug-drug interaction (DDI). Metode SLR dipilih karena mampu memberikan analisis yang

komprehensif, sistematis, dan dapat direplikasi terhadap perkembangan penelitian dalam suatu bidang [1]. Proses pelaksanaan review mengikuti pedoman **PRISMA 2020**, yang menjamin transparansi dalam proses seleksi studi, ekstraksi data, dan sintesis hasil sehingga meningkatkan validitas dan reliabilitas temuan.

Tahapan utama dalam SLR ini meliputi: (1) perumusan pertanyaan penelitian, (2) pengembangan strategi pencarian literatur, (3) seleksi studi berdasarkan kriteria inklusi dan eksklusi, (4) ekstraksi data, (5) penilaian kualitas studi, serta (6) sintesis dan analisis data. Seluruh tahapan didokumentasikan secara sistematis untuk memastikan reproducibility dan memungkinkan pembaruan studi di masa mendatang.

### 2.1.1 Sumber Database

Pencarian literatur dilakukan pada lima database utama yang mencakup bidang bioinformatika, farmasi, dan ilmu komputer, yaitu PubMed/PMC, IEEE Xplore, Scopus, Web of Science, dan Google Scholar. Penggunaan beberapa database bertujuan untuk memaksimalkan cakupan literatur serta meminimalkan bias publikasi [3][4].

Seluruh hasil pencarian dari berbagai database digabungkan dan dilakukan proses deduplikasi menggunakan reference management software untuk memastikan tidak ada duplikasi data.

### 2.1.2 Kata Kunci dan String Pencarian

Strategi pencarian dikembangkan berdasarkan tiga konsep utama, yaitu: (1) drug-drug interaction, (2) machine learning, dan (3) prediksi atau klasifikasi. Kata kunci dikombinasikan menggunakan operator Boolean (AND, OR) untuk menghasilkan pencarian yang seimbang antara sensitivitas dan spesifisitas.

Pendekatan ini memastikan bahwa studi yang relevan dapat teridentifikasi secara optimal, termasuk penelitian yang menggunakan berbagai algoritma seperti Support Vector Machine, Random Forest, hingga deep learning dan Graph Neural Networks [2][4].

### 2.1.3 Periode Pencarian

Literatur yang dikaji dibatasi pada periode **2019–2025**, dengan pertimbangan bahwa periode ini mencerminkan perkembangan terbaru dalam metode machine learning, khususnya deep learning dan pendekatan berbasis graf dalam prediksi DDI [2][7]. Batasan ini juga bertujuan untuk memastikan relevansi metodologi dengan kondisi state-of-the-art saat ini.

### 2.1.4 Pencarian Tambahan

Selain pencarian database, dilakukan juga strategi tambahan seperti:

- penelusuran referensi (backward citation),
- pencarian sitasi ke depan (forward citation),
- eksplorasi grey literature seperti arXiv,
- serta penelusuran manual pada proceeding konferensi utama.

Pendekatan ini penting untuk memastikan tidak ada studi relevan yang terlewat, terutama penelitian terbaru yang belum terindeks secara formal [1].

## 2.2 Kriteria Seleksi Studi

### 2.2.1 Kriteria Inklusi

Studi yang diikutsertakan harus memenuhi kriteria berikut:

1. Berfokus pada prediksi DDI menggunakan metode machine learning atau deep learning.
2. Merupakan penelitian asli (original research) yang dipublikasikan pada jurnal atau konferensi bereputasi.
3. Menggunakan dataset empiris serta melaporkan metrik evaluasi kuantitatif seperti accuracy, precision, recall, F1-score, atau AUC.
4. Menyediakan deskripsi metodologi yang jelas dan dapat direproduksi.
5. Dipublikasikan dalam bahasa Inggris pada periode 2019–2025.

Kriteria ini dirancang untuk memastikan kualitas dan relevansi studi yang dianalisis [4].

### 2.2.2 Kriteria Eksklusi

Studi dikeluarkan apabila:

- bukan penelitian utama (misalnya review atau editorial),
- tidak berfokus pada DDI (misalnya hanya drug-target interaction),
- tidak menggunakan pendekatan machine learning,
- tidak memiliki evaluasi kuantitatif,
- atau memiliki metodologi yang tidak jelas.

Langkah ini bertujuan menjaga konsistensi dan validitas hasil analisis.

## 2.3 Proses Seleksi dan Screening

Seleksi studi dilakukan berdasarkan alur PRISMA yang terdiri dari empat tahap: identifikasi, deduplikasi, screening, dan evaluasi full-text.

Proses awal menghasilkan lebih dari seribu artikel dari berbagai database, yang kemudian dikurangi melalui proses deduplikasi. Selanjutnya dilakukan screening berdasarkan judul dan abstrak oleh dua reviewer independen untuk menentukan relevansi awal.

Tahap berikutnya adalah evaluasi full-text terhadap studi yang lolos screening awal. Studi yang tidak memenuhi kriteria metodologi atau tidak melaporkan hasil evaluasi kuantitatif dikeluarkan. Proses ini menghasilkan sejumlah studi akhir yang digunakan dalam analisis.

Pendekatan multi-reviewer digunakan untuk meningkatkan objektivitas, dengan pengukuran kesepakatan menggunakan koefisien statistik seperti Cohen's Kappa [3].

## 2.4 Ekstraksi Data

Data dari setiap studi diekstrak menggunakan form terstruktur yang mencakup beberapa aspek utama, yaitu:

- informasi bibliografis,
- karakteristik studi,
- jenis algoritma machine learning,
- representasi fitur,
- dataset yang digunakan,
- metodologi evaluasi,

- hasil performa,
- serta aspek interpretabilitas model.

Ekstraksi dilakukan oleh dua reviewer secara independen untuk meminimalkan bias, dan perbedaan hasil diselesaikan melalui diskusi.

Pendekatan ini memungkinkan analisis yang sistematis terhadap berbagai metode ML yang digunakan dalam prediksi DDI, termasuk perbandingan performa antar algoritma [2][4].

## 2.5 Penilaian Kualitas Studi

Kualitas metodologi setiap studi dinilai menggunakan kriteria yang diadaptasi dari **PROBAST**, yang mencakup aspek seperti:

- kejelasan tujuan penelitian,
- kualitas dataset,
- transparansi model,
- strategi validasi,
- serta kelengkapan metrik evaluasi.

Setiap studi diberikan skor kualitas yang kemudian diklasifikasikan menjadi tinggi, sedang, atau rendah. Penilaian ini penting untuk memahami reliabilitas hasil dan mengidentifikasi potensi bias dalam penelitian [1].

Selain itu, dilakukan juga analisis risiko bias yang mencakup:

- bias publikasi,
- bias seleksi dataset,
- data leakage,
- serta bias pelaporan hasil.

Evaluasi ini membantu memastikan bahwa hasil review tidak terdistorsi oleh kelemahan metodologis dari studi yang dianalisis.

## 2.6 Sintesis dan Analisis Data

### 2.6.1 Sintesis Kualitatif

Sintesis kualitatif dilakukan untuk:

- menggambarkan perkembangan metode ML dari waktu ke waktu,
- mengidentifikasi tren penggunaan algoritma,

- serta merangkum tantangan dan arah penelitian ke depan.

Perkembangan signifikan terlihat pada pergeseran dari metode tradisional ke deep learning dan graph-based approaches yang lebih kompleks dan akurat [2][4].

### 2.6.2 Sintesis Kuantitatif

Analisis kuantitatif meliputi:

- distribusi frekuensi algoritma dan dataset,
- perbandingan performa antar metode,
- serta analisis statistik terhadap metrik evaluasi.

Jika memungkinkan, dilakukan meta-analisis untuk memperoleh estimasi performa gabungan dan mengukur heterogenitas antar studi. Pendekatan ini memberikan gambaran komprehensif mengenai efektivitas berbagai metode ML dalam prediksi DDI [3].

### 2.6.3 Tools dan Software

Analisis data dilakukan menggunakan berbagai tools seperti Microsoft Excel untuk manajemen data, R dan Python untuk analisis statistik dan visualisasi, serta software manajemen referensi untuk pengolahan literatur.

Penggunaan tools ini mendukung proses analisis yang sistematis, terstruktur, dan dapat direplikasi.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Karakteristik Studi

Hasil pencarian literatur menghasilkan 1.285 publikasi, yang setelah proses seleksi berbasis PRISMA menghasilkan 60 studi yang memenuhi kriteria inklusi. Sebagian besar studi yang dieliminasi disebabkan oleh kurangnya detail metodologi, tidak adanya evaluasi kuantitatif, atau fokus pada drug-target interaction (DTI) alih-alih drug-drug interaction (DDI). Hal ini menunjukkan belum adanya standardisasi pelaporan dalam penelitian machine learning (ML) di bidang farmasi [1][4].

Distribusi publikasi menunjukkan tren peningkatan signifikan pada periode 2019–2023, mencerminkan berkembangnya minat

terhadap penerapan ML dalam prediksi DDI, terutama sejak kemunculan deep learning dan graph-based methods [2][4]. Secara geografis, penelitian didominasi oleh negara Asia, khususnya China, diikuti Amerika Utara dan Eropa, yang menunjukkan konsentrasi riset AI di kawasan tersebut.

Mayoritas studi (63,3%) memiliki kualitas tinggi berdasarkan penilaian metodologi, namun masih ditemukan korelasi antara kualitas studi dengan performa yang dilaporkan, yang mengindikasikan potensi bias dalam pelaporan hasil [3].

### 3.2 Algoritma Machine Learning untuk Prediksi DDI

Empat kategori utama algoritma diidentifikasi: traditional ML (25%), deep learning (30%), graph-based methods (36,7%), dan hybrid/ensemble (8,3%).

Metode berbasis graf menjadi pendekatan paling dominan dan menunjukkan performa terbaik karena kemampuannya merepresentasikan struktur molekul dan hubungan biologis secara alami [2][8]. Secara umum, terjadi evolusi metodologi sebagai berikut:

- **Traditional ML** (SVM, Random Forest) digunakan pada tahap awal dengan performa moderat (F1 ~0,84–0,87) dan keunggulan pada interpretabilitas serta efisiensi komputasi.
- **Deep Learning** (CNN, RNN, DNN) meningkatkan performa (F1 ~0,89–0,92) melalui pembelajaran fitur otomatis, namun memiliki keterbatasan pada interpretabilitas dan kebutuhan komputasi tinggi [4].
- **Graph Neural Networks (GNN)** menjadi state-of-the-art dengan performa tertinggi (F1 ~0,93–0,96) karena mampu menangkap relasi struktural kompleks [2].
- **Hybrid/Multi-modal models** mencapai performa terbaik (F1 hingga ~0,98) dengan menggabungkan berbagai sumber data [8].

Analisis statistik menunjukkan perbedaan signifikan antar kategori algoritma ( $p < 0,001$ ), dengan urutan performa: Hybrid > Graph-based > Deep Learning > Traditional ML.

### 3.3 Representasi Fitur dan Data

Representasi data memainkan peran penting dalam performa model. Tiga pendekatan utama adalah:

1. **Struktur kimia** (SMILES, fingerprint, molecular graph)
  - Graph representation memberikan performa terbaik karena mempertahankan informasi struktur molekul.
2. **Similarity-based features**
  - Menggunakan kemiripan struktur, efek samping, atau target protein untuk meningkatkan prediksi.
3. **Network-based representation**
  - Menggunakan embedding dari knowledge graph atau jaringan biologis.

Pendekatan **multi-modal** yang menggabungkan berbagai jenis data memberikan performa tertinggi (F1 ~0,945), namun dengan biaya komputasi yang jauh lebih tinggi. Hal ini menunjukkan adanya trade-off antara akurasi dan kompleksitas model [2][4].

### 3.4 Dataset dan Benchmark

Dataset yang paling umum digunakan adalah DrugBank (63,3%), diikuti TWOSIDES dan BIOSNAP. Dataset ini memiliki karakteristik berbeda, namun semuanya menghadapi masalah utama berupa **class imbalance** (rasio positif:negatif hingga 1:50) [3].

Strategi penanganan imbalance seperti class weighting dan focal loss terbukti meningkatkan performa hingga 5–9%. Namun, tidak semua studi mengimplementasikan teknik ini, yang dapat menyebabkan bias evaluasi.

Metode validasi yang digunakan bervariasi, dengan k-fold cross-validation sebagai pendekatan paling umum. Namun, validasi temporal dan skenario cold-start menunjukkan penurunan performa signifikan (8–20%), yang mengindikasikan bahwa banyak studi masih menghasilkan estimasi performa yang terlalu optimis [4].

### 3.5 Performa Model

Secara umum, performa model ML dalam prediksi DDI cukup tinggi, dengan rata-rata:

- F1-score: ~0,90
- AUC-ROC: ~0,93

Model terbaik (2023–2025) mencapai F1 hingga ~0,98, didominasi oleh pendekatan multi-modal berbasis graph [8]. Namun, terdapat indikasi adanya **publication bias**, di mana hasil dengan performa tinggi lebih sering dilaporkan.

Analisis menunjukkan bahwa ukuran dataset berpengaruh signifikan terhadap performa ( $R^2 = 0,68$ ), dengan peningkatan performa seiring bertambahnya data, meskipun terdapat *diminishing returns* pada dataset besar. Pada dataset kecil, metode tradisional masih kompetitif karena lebih stabil dan efisien.

### 3.6 Tantangan dan Limitasi

Beberapa tantangan utama yang diidentifikasi meliputi:

#### 1. Data-related challenges

- **Data scarcity:** hanya sebagian kecil pasangan obat memiliki data interaksi
- **Class imbalance:** menyebabkan bias model
- **Noise dan inkonsistensi data:** terutama pada dataset berbasis laporan klinis

#### 2. Cold-start problem

Model kesulitan memprediksi interaksi untuk obat baru, dengan penurunan performa hingga ~18% [4].

#### 3. Interpretabilitas model

Sebanyak 45% studi menggunakan model black-box tanpa penjelasan, yang menjadi hambatan dalam adopsi klinis. Pendekatan explainable AI mulai digunakan, seperti attention dan SHAP, namun masih terbatas [8].

#### 4. Kompleksitas komputasi

Model modern memerlukan sumber daya besar, hingga 10–30 kali lebih tinggi dibanding metode

tradisional, sehingga membatasi implementasi praktis.

#### 5. Kurangnya standardisasi

Variasi dataset, metode evaluasi, dan metrik menyebabkan sulitnya perbandingan antar studi. Selain itu, hanya sekitar 11,7% studi yang benar-benar reproducible.

### 3.7 Tren dan Arah Penelitian Masa Depan

Beberapa tren utama yang berkembang meliputi:

- **Explainable AI (XAI)** untuk meningkatkan kepercayaan klinis
- **Foundation models dan transfer learning** untuk meningkatkan generalisasi
- **Multi-modal learning** untuk integrasi data heterogen
- **Integrasi dengan Electronic Health Records (EHR)** untuk aplikasi dunia nyata [3]
- **Mechanism-aware prediction** untuk memahami jenis interaksi
- **Polypharmacy modeling** untuk interaksi lebih dari dua obat
- **Federated learning** untuk menjaga privasi data pasien

Pendekatan ini diharapkan dapat menjembatani kesenjangan antara performa model di laboratorium dan aplikasi klinis nyata.

### 3.8 Sintesis dan Analisis Kritis

Secara keseluruhan, **Graph Neural Networks** merupakan metode paling efektif untuk prediksi DDI, terutama pada dataset besar, sementara pendekatan multi-modal memberikan peningkatan tambahan dengan biaya komputasi tinggi [2][8].

Pemilihan metode terbaik bergantung pada:

- ukuran dataset
- kebutuhan interpretabilitas
- ketersediaan sumber daya komputasi

Selain itu, strategi validasi yang tepat (termasuk temporal dan cold-start) sangat penting untuk mendapatkan estimasi performa yang realistis.

### 3.9 Implikasi

#### 1. Untuk Peneliti

- Perlu standarisasi benchmark dan evaluasi
- Fokus pada cold-start dan data dunia nyata
- Pengembangan model yang interpretable

#### 2. Untuk Industri dan Klinisi

- ML sebaiknya digunakan sebagai decision support, bukan pengganti keputusan klinis
- Integrasi dengan workflow klinis sangat penting
- Perlu manajemen false positive untuk menghindari alert fatigue

#### 3. Untuk Regulator

- Diperlukan framework validasi AI dalam kesehatan
- Standarisasi dan transparansi model
- Monitoring performa pasca implementasi

Sintesis data dilakukan secara kualitatif dan kuantitatif. Analisis kualitatif digunakan untuk mengidentifikasi tren perkembangan metode ML dalam prediksi DDI, sementara analisis kuantitatif digunakan untuk membandingkan performa berbagai algoritma berdasarkan metrik evaluasi standar. Pendekatan ini memungkinkan identifikasi metode yang paling efektif serta faktor-faktor yang mempengaruhi performa model, seperti jenis dataset dan representasi fitur [2][4].

## 4. KESIMPULAN DAN SARAN

### 4.1 Kesimpulan

Penelitian ini menunjukkan bahwa **machine learning, khususnya Graph Neural Networks (GNN)**, merupakan pendekatan paling efektif untuk prediksi *drug-drug interaction (DDI)*, dengan performa tertinggi dibanding metode lain. Pendekatan **multi-modal** (menggabungkan struktur kimia, data biologis, dan fenotipik) mampu memberikan akurasi terbaik, meskipun dengan biaya komputasi yang jauh lebih tinggi.

Pemilihan metode sangat bergantung pada kondisi:

- **Dataset kecil** → metode tradisional (Random Forest, SVM) masih kompetitif
- **Dataset menengah-besar** → GNN menjadi pilihan utama
- **Kebutuhan akurasi tinggi** → hybrid/multi-modal terbaik

Dalam hal representasi data, **molecular graph** paling optimal untuk menangkap struktur dan relasi obat, sementara kombinasi multi-modal memberikan hasil terbaik secara keseluruhan.

Namun, terdapat beberapa tantangan utama:

1. **Cold-start problem** (obat baru sulit diprediksi)
2. **Ketidakeimbangan data** (interaksi positif sangat sedikit)
3. **Kurangnya interpretabilitas model**
4. **Gap antara performa laboratorium dan dunia klinis**
5. **Tidak adanya standar evaluasi dan rendahnya reproducibility**
6. **Keterbatasan pada interaksi lebih dari dua obat (polypharmacy)**

Selain itu, performa yang dilaporkan dalam literatur cenderung **lebih optimis** dibanding kondisi nyata, akibat bias publikasi dan metode validasi yang kurang realistis.

### 4.2 Saran

#### 4.2.1 Untuk Peneliti:

Prioritas penelitian masa depan sebaiknya fokus pada: (1) solusi *cold-start problem* melalui *meta-learning*, *transfer learning* dari *pre-trained molecular models*, dan *few-shot learning* untuk prediksi dengan data minimal; (2) pengembangan *explainable AI* yang dapat memvalidasi prediksi terhadap mekanisme farmakologi yang diketahui, menggunakan *attention visualization*, *SHAP values*, atau *subgraph explanation*; (3) validasi eksternal multi-pusat pada data EHR dari berbagai institusi untuk menilai generalisasi; (4) integrasi faktor pasien-spesifik (usia, genetika, fungsi organ) menuju prediksi DDI yang dipersonalisasi; dan (5) adopsi praktik *open science* dengan *mandatory code/data sharing* dan penggunaan dataset *pre-split* untuk standarisasi evaluasi.

**Pemilihan metode sebaiknya disesuaikan dengan skenario:** untuk dataset besar (>50K pairs) gunakan GNN atau *multi-modal approaches*; untuk dataset kecil (<10K pairs) *Random Forest* masih kompetitif dengan kompleksitas lebih rendah; untuk kebutuhan interpretabilitas tinggi, pertimbangkan *logistic regression* dengan *molecular descriptors*; dan untuk aplikasi *real-time*, pertimbangkan *trade-off* antara akurasi dan *computational cost*.

#### 4.2.2 Untuk Industri Farmasi:

Implementasi sistem prediksi DDI berbasis ML memerlukan pendekatan bertahap: dimulai dengan deployment *shadow mode* untuk validasi performa tanpa mempengaruhi proses pengembangan obat, diikuti integrasi ke *early-stage screening* untuk memprioritaskan eksperimen *in vitro*, dan ekspansi ke *clinical trial planning* dengan *continuous monitoring*. *Threshold* keputusan harus dikalibrasi berdasarkan tingkat risiko: *high recall* (>0,95) untuk interaksi berpotensi fatal meski dengan *false positives* lebih tinggi, dan *high precision* (>0,90) untuk screening skala besar guna menghindari eliminasi kandidat obat yang sebenarnya aman.

#### 4.2.3 Untuk Pembuat Kebijakan dan Regulator:

Pengembangan framework regulatori untuk validasi sistem ML dalam prediksi DDI sangat diperlukan, mencakup: standar validasi dengan persyaratan *external validation* pada minimal dua dataset independen, *mandatory reporting* performa pada skenario *cold-start* dan *temporal validation*, dan dokumentasi transparansi model termasuk arsitektur, data pelatihan, dan limitasi yang diketahui. Klasifikasi berbasis risiko dapat membedakan persyaratan antara sistem untuk *early-stage screening* (persyaratan lebih ringan) versus *clinical decision support* (validasi ketat).

## 5. PENUTUP

Interaksi obat merupakan tantangan penting dalam farmakoterapi, dan **machine learning (terutama GNN)** telah menunjukkan performa tinggi dalam prediksi DDI. Namun, keberhasilan komputasional belum otomatis berdampak klinis.

Masih diperlukan solusi untuk masalah utama seperti **generalisasi obat baru, interpretabilitas, validasi dunia nyata, dan penerapan yang bertanggung jawab**, yang menuntut kolaborasi lintas disiplin.

Ke depan, fokus utama adalah **standardisasi evaluasi, peningkatan reproducibility, validasi klinis, explainability, dan implementasi berkelanjutan**.

Secara keseluruhan, bidang ini berada pada fase transisi menuju penerapan nyata, dengan tujuan akhir **meningkatkan keamanan dan efektivitas penggunaan obat bagi pasien**.

## PERNYATAAN PENGHARGAAN

Terima kasih penulis sampaikan kepada Institut Teknologi dan Kesehatan Bintang Persada atas fasilitas yang diberikan untuk melakukan penelitian ini.

## DAFTAR PUSTAKA

- [1] Gheorghita, F.-I., Bocanet, V.-I., & Iantovics, L. B. (2025). Machine learning-based drug-drug interaction prediction: A critical review of models, limitations, and data challenges. *Frontiers in Pharmacology, 16*, 1632775. <https://doi.org/10.3389/fphar.2025.1632775>
- [2] Li, S. (2024). Deep learning for drug-drug interaction prediction: A comprehensive review. *Quantitative Biology, 12*(1), 30–52. <https://doi.org/10.1002/qub2.32>
- [3] Hu, Q., Chen, Y., Zou, D., He, Z., & Xu, T. (2024). Predicting adverse drug event using machine learning based on electronic health records: A systematic review and meta-analysis. *Frontiers in Pharmacology, 15*, 1497397. <https://doi.org/10.3389/fphar.2024.1497397>
- [4] Han, K., Cao, P., Wang, Y., Xie, F., Ma, J., Yu, M., Wang, J., Xu, Y., Zhang, Y., & Wan, J. (2022). A review of approaches for predicting drug–drug interactions based on machine learning. *Frontiers in Pharmacology, 12*, 814858. <https://doi.org/10.3389/fphar.2021.814858>
- [5] Drug-Target Interaction/Affinity Prediction: Deep Learning Models and Advances Review. (2025). *arXiv preprint*. <https://arxiv.org/pdf/2502.15346>

- [6] Mei, S., & Zhang, K. (2021). A machine learning framework for predicting drug–drug interactions. *Scientific Reports*, *11*, 17619. <https://doi.org/10.1038/s41598-021-97193-8>
- [7] Lee, G., Park, C., & Ahn, J. (2019). Novel deep learning model for more accurate prediction of drug-drug interaction effects. *BMC Bioinformatics*, *20*, 415. <https://doi.org/10.1186/s12859-019-3013-0>
- [8] Li, R., Zhang, Y., Sun, H., Lin, S., Jia, G., Fang, Y., Zhang, C., Song, X., Zhao, J., Hu, L., Yuan, Y., Mao, X., Li, J., Kaushik, A., An, D., & Wei, D. (2025). Towards interpretable drug interaction prediction via dual-stage attention and Bayesian calibration with active learning. *PeerJ Computer Science*, *11*, e2847. <https://doi.org/10.7717/peerj-cs.2847>
- [9] Huang, K., Fu, T., Glass, L., Zitnik, M., Xiao, C., & Sun, J. (2020). DeepPurpose: A deep learning library for drug-target interaction prediction and applications to repurposing and screening. *arXiv preprint*. <https://arxiv.org/pdf/2004.08919>
- [10] Biomolecular interaction prediction: The era of AI. (2024). *Preprints*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12520538/>
- [11] Ryu, J. Y., Kim, H. U., & Lee, S. Y. (2019). Deep learning for high-order drug-drug interaction prediction. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* (pp. 662–667). Association for Computing Machinery. <https://doi.org/10.1145/3307339.3342136>
- [12] Chaudhari, S., Khemani, B., Patil, S., & Gupta, J. (2024). Genome sequence analysis and drug-target interaction prediction using deep learning. In A. E. Hassanien, S. Anand, A. Jaiswal, & P. Kumar (Eds.), *Innovative Computing and Communications. ICICC 2024. Lecture Notes in Networks and Systems* (Vol. 1038, pp. 545–558). Springer. [https://doi.org/10.1007/978-981-97-4149-6\\_39](https://doi.org/10.1007/978-981-97-4149-6_39)
- [13] A hybrid computational intelligence framework with metaheuristic optimization for drug-drug interaction prediction. (2024). *arXiv preprint*. <https://arxiv.org/pdf/2510.09668>