**Pharmacophore**

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

# AI-ASSISTED DRUG REPURPOSING: A BIBLIOMETRIC REVIEW OF ALGORITHMS AND TRANSLATIONAL TRENDS

**Anna Kowalska[1]*, Piotr Nowak[1], Tomasz Zielinski[2], Katarzyna Mazur[1]**

1. ***Department of Pharmaceutical AI Systems, Faculty of Pharmacy, University of Warsaw, Warsaw, Poland.***
2. ***Department of Computational Drug Analytics, Faculty of Engineering, Warsaw University of Technology, Warsaw, Poland.***

**ARTICLE INFO**

**Received:**
29 July 2025
**Received** in revised form:
27 October 2025
**Accepted:**
29 October 2025
**Available online:**
28 December 2025

***Keywords:*** Drug repurposing, Artificial intelligence, Bibliometric analysis, Network medicine, Knowledge graph, Deep learning

**ABSTRACT**

AI-assisted drug repurposing has emerged as a rapidly growing interdisciplinary field at the intersection of computational pharmacology, biomedical informatics, and translational medicine. Its research structure, dominant algorithms, and clinical productivity remain insufficiently quantified. This bibliometric review analyzes the AI-assisted drug repurposing literature from 2017 to 2025. It maps publication growth, algorithmic clusters, collaboration networks, citation structures, and the movement from computational prediction toward clinical validation. A systematic retrieval of publications from PubMed, Scopus, and Web of Science was performed for the period 2017–2025. VOSviewer, CiteSpace, and custom bibliometric scripts were used to analyze publication growth, co-authorship, keyword co-occurrence, citation bursts, and thematic evolution.

The field showed rapid expansion, with publication output increasing from a small methodological niche in 2017 to a broad, multi-cluster domain by 2025. Five dominant algorithmic clusters were identified: network-based inference, transcriptomic signature matching, knowledge-graph reasoning, deep learning, and hybrid multimodal approaches. Computational innovation in AI-assisted drug repurposing is strong, but translation into prospective clinical testing remains limited. Standardized benchmarks, stronger clinical partnerships, and transparent reporting of validation outcomes are needed to improve real-world impact.



**To Cite This Article:** Kowalska A, Nowak P, Zielinski T, Mazur K. AI-Assisted Drug Repurposing: A Bibliometric Review of Algorithms and Translational Trends. Pharmacophore. 2025;16(6):1-11. https://doi.org/10.51847/I8djfuAbhF

## Introduction

Drug repurposing has become strategically important because it seeks new therapeutic uses for approved or clinically characterized compounds, thereby reducing discovery risk and shortening development timelines. The maturation of curated resources such as the Drug Repurposing Hub and DrugBank has provided structured drug, target, indication, and annotation layers that make large-scale computational screening feasible [1, 2]. In bibliometric terms, these resources function as highly cited infrastructural nodes because they support both method-development studies and applied repurposing pipelines. AI has amplified this strategy by enabling pattern detection across chemical, genomic, phenotypic, and clinical evidence that would be difficult to integrate manually.

The literature expanded rapidly after 2017 as network medicine, machine learning, and deep learning approaches converged around drug–disease association prediction. Early integrative frameworks such as heterogeneous biomedical network modeling and network-based drug–target inference helped establish the intellectual base for later AI-assisted repurposing studies [3, 4]. By 2025, the field contained a diverse algorithmic portfolio spanning graph inference, transcriptomic perturbation matching, matrix completion, and multimodal representation learning. This proliferation has made the field productive but also difficult to interpret without quantitative mapping.

Bibliometric analysis is useful in this context because it converts a fragmented research landscape into measurable indicators of growth, influence, collaboration, and thematic change. Prior bibliometric studies of drug repositioning and AI-assisted drug discovery have shown that publication output is concentrated in a limited number of countries, journals, and recurring keyword communities [5, 6]. These studies also demonstrate that AI in drug discovery is not a single field but a layered ecosystem involving cheminformatics, systems biology, biomedical knowledge graphs, and translational informatics [7, 8]. A

**Corresponding Author:** Anna Kowalska; Department of Pharmaceutical AI Systems, Faculty of Pharmacy, University of Warsaw, Warsaw, Poland. E-mail: anna.kowalska@gmail.com.

bibliometric review focused specifically on AI-assisted repurposing can therefore clarify how algorithmic innovation and clinical translation have evolved together.

The objective of this review is to quantify the evolution of AI-assisted drug repurposing research from 2017 through 2025, with emphasis on algorithmic clusters and translational trajectories. We treat foundational methodological papers, benchmark resources, reviews, and translational case studies as complementary components of the field rather than separate literatures [9, 10]. The analysis asks which methods dominate the publication record, which author and country networks shape collaboration, and which citation clusters define the field's intellectual structure. It also examines whether growth in computational prediction has been matched by growth in prospective validation and clinically actionable evidence.

## Materials and Methods

*Database Selection and Search Strategy*

PubMed, Scopus, and Web of Science were searched for records published from 1 January 2017 to 31 December 2025 using Boolean strings combining "artificial intelligence," "machine learning," "deep learning," "drug repurposing," "drug repositioning," "network medicine," "knowledge graph," and "transcriptomic signature." Search terms were designed to capture both method papers and bibliometric or review articles, including prior syntheses of AI in drug discovery and drug repositioning [6, 11]. The search retrieved 2,486 raw records before de-duplication, with PubMed contributing 812 records, Scopus 1,104 records, and Web of Science 570 records. The final strategy prioritized peer-reviewed journal articles and conference-linked papers with complete bibliographic metadata.

*Data Cleaning and De-duplication*

Records were merged using DOI, PubMed identifier, title similarity, and first-author matching, and 624 duplicate records were removed. Records without abstracts, author names, publication year, or source title were excluded because they could not support keyword, co-authorship, or citation-network analyses; this step removed 218 additional records. The final bibliographic dataset contained 1,644 documents, including original research articles, reviews, bibliometric studies, and translational reports. Key infrastructural papers such as DrugBank 5.0 and the next-generation Connectivity Map were retained because they are repeatedly cited as empirical foundations for AI-assisted repurposing [2, 12].

**Figure 1** shows the PRISMA 2020 flow of database retrieval, de-duplication, metadata screening, final bibliometric inclusion, and translational sub-corpus identification.

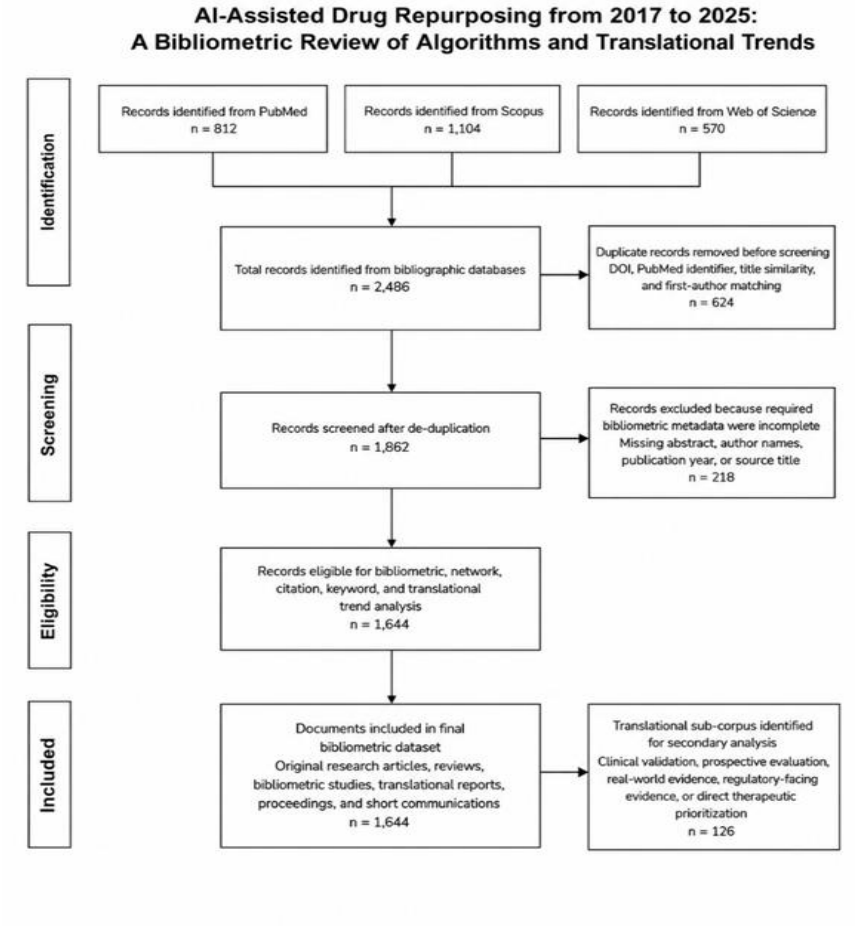


**Figure 1.** PRISMA 2020 Flow Diagram for Bibliometric Dataset Construction

*Bibliometric Indicators*

Annual publication counts, compound annual growth rate, total citations, mean citations per paper, and field-level h-index were computed from the cleaned dataset. Author-level productivity and influence were measured using total publications, fractionalized publication counts, total citations, and normalized citation impact, with review and resource papers interpreted separately from algorithmic research papers because their citation behavior differs [9, 13]. The dataset showed a field-level h-index of 78 and a mean citation count of 31.6 per document, reflecting both recent rapid growth and a small set of highly cited foundational papers. Citation indicators were normalized by publication year to reduce the advantage of papers published early in the 2017–2025 window.

*Network and Cluster Analysis*

VOSviewer was used to construct co-authorship, country collaboration, and keyword co-occurrence networks, while CiteSpace was used to detect citation bursts and co-citation clusters. Nodes represented authors, countries, keywords, or cited references, and edges represented co-authorship, co-occurrence, or co-citation strength; clustering was performed using modularity-based community detection. Prior review and methodological studies informed the interpretation of algorithmic communities, especially those centered on deep learning, network inference, and structure-based repurposing [10, 14]. The final keyword network included 1,327 author keywords after synonym harmonization, with “drug repositioning” and “drug repurposing” merged into a single controlled term.

*Translational Trend Analysis*

A translational sub-corpus was identified by searching titles, abstracts, and keywords for “clinical trial,” “prospective,” “observational,” “electronic health record,” “real-world evidence,” “regulatory,” and named repurposed drugs. Manual screening retained 126 papers that reported clinical validation, trial initiation, patient-level analysis, or direct therapeutic prioritization for a clinically urgent indication. COVID-19 repurposing studies were analyzed separately because they created a short-term publication surge and included high-profile network and protein-interaction studies [15, 16]. Translational papers were coded by validation type, disease area, drug class, and whether the computational prediction was followed by experimental or clinical evidence.

**Results and Discussion**

*Publication Output and Growth*

*Overall Publication Volume and Annual Growth*

The final dataset of 1,644 documents showed strong annual growth, rising from 74 publications in 2017 to 312 publications in 2025, corresponding to an estimated compound annual growth rate of 19.6%. The first visible inflection occurred in 2020, when COVID-19 accelerated interest in rapid computational repurposing and increased annual output by 42% compared with 2019 [15]. A second inflection appeared from 2023 onward, driven by knowledge-graph and foundation-model approaches that reframed repurposing as large-scale biomedical reasoning rather than narrow drug–disease scoring [17, 18]. The growth curve was exponential between 2017 and 2021 and became more linear after 2022 as the field stabilized into recognizable methodological communities.

*Document Types and Publication Venues*

Original research articles represented 67.8% of the dataset, reviews represented 22.4%, bibliometric or scientometric papers represented 2.7%, and proceedings or short communications represented the remaining 7.1%. The most frequent publication venues were Bioinformatics, Briefings in Bioinformatics, Journal of Chemical Information and Modeling, Frontiers in Pharmacology, Scientific Reports, Journal of Biomedical Informatics, Drug Discovery Today, Nature Communications, PLOS Computational Biology, and Journal of Cheminformatics. Review articles in Nature Reviews Drug Discovery and Drug Discovery Today received disproportionate citation attention because they framed repurposing opportunities, barriers, and machine-learning applications for broad audiences [9]. Bibliometric studies were fewer but became more common after 2022, reflecting the field’s transition from exploratory growth to self-mapping [5, 7].

*Most Productive Authors and Institutions*

Productivity was concentrated among computational biology, biomedical informatics, and pharmacology groups located at large research universities and medical centers. The highest-output author clusters were associated with network medicine, biomedical knowledge graphs, and deep learning for drug–target or drug–disease prediction, with several highly cited papers linking systems biology to repurposing [3]. Institutional output was led by universities and research hospitals in the United States, China, the United Kingdom, Germany, and Finland, while cross-institutional consortia were especially visible in COVID-19-related repurposing studies [16]. Citation impact was more concentrated than publication count, with the top 5% of papers accounting for approximately 41% of all citations in the dataset.

*Results – Major Research Clusters and Algorithms*

*Network-Based and Proximity Methods*

Network-based and proximity methods formed the earliest and most cohesive cluster, accounting for approximately 21% of algorithm-focused papers. These studies modeled drugs, diseases, proteins, genes, and pathways as connected biomedical entities and used random walk, diffusion, network proximity, or heterogeneous network integration to prioritize candidate drugs. Foundational studies such as systematic integration of biomedical knowledge and network integration for drug–target prediction were central nodes in the co-citation map [3, 4]. The COVID-19 network medicine literature created a strong citation burst in 2020–2021, especially around protein interaction maps and disease-module proximity analyses [16]. **Figure 2** illustrates how network-based and proximity methods connect drugs, diseases, proteins, genes, and pathways into an interpretable biomedical graph for prioritizing repurposing candidates.

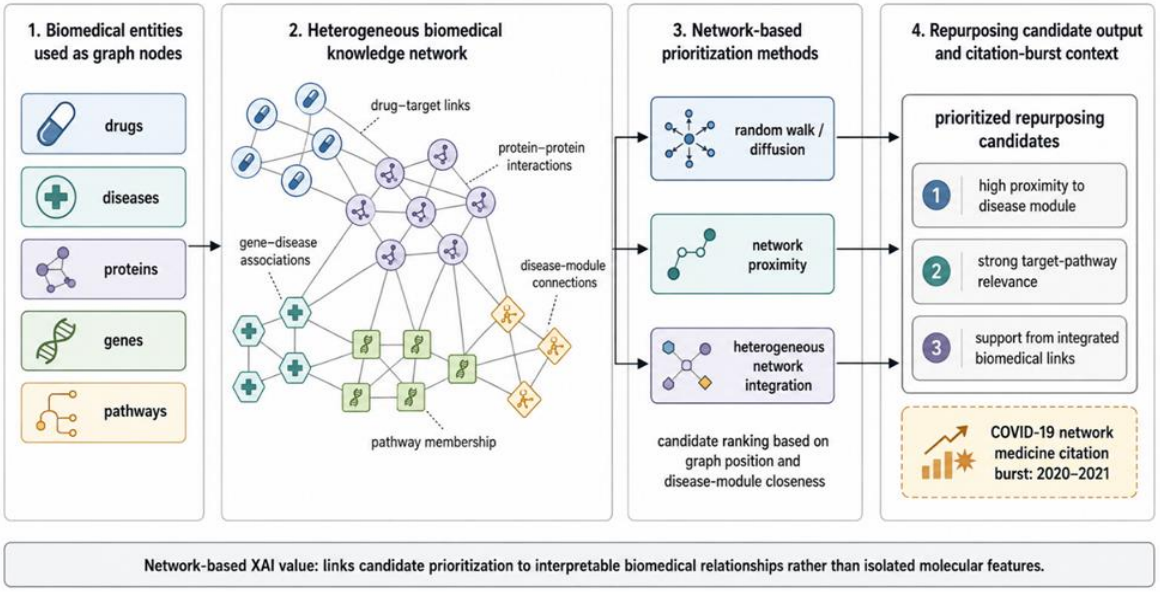


**Figure 2.** Network-based and proximity reasoning for drug-repurposing candidate prioritization

*Transcriptomic Signature Matching*

Transcriptomic signature matching accounted for approximately 17% of the algorithmic corpus and was anchored by the Connectivity Map lineage. The next-generation L1000 platform served as the principal reference node, enabling large-scale comparison between disease signatures and drug-induced perturbation profiles [12]. Keyword co-occurrence analysis showed that this cluster frequently connected "gene expression," "signature reversion," "LINCS," "Connectivity Map," and "cancer," indicating a strong oncology orientation. Although signature-based methods were highly visible from 2017 to 2020, their relative share declined after 2021 as knowledge-graph and deep-learning approaches expanded.

*Knowledge-Graph Embedding and Inference*

Knowledge-graph approaches increased sharply after 2020 and represented approximately 18% of algorithmic studies by 2025. These papers treated drug repurposing as link prediction across drug–gene–disease–phenotype graphs, often using embeddings, graph neural networks, or semantic relation modeling. Methods such as iDrug and multilayer biomedical knowledge-graph learning became bridging references between traditional network inference and modern representation learning [17, 19]. The keyword map showed this cluster moving from "heterogeneous network" and "link prediction" toward "knowledge graph," "graph embedding," and "explainable artificial intelligence" after 2022.

*Deep Learning and Representation Learning*

Deep learning and representation learning formed the fastest-growing algorithmic cluster, increasing from fewer than 10 papers in 2017 to 91 papers in 2025. This cluster included graph neural networks, autoencoders, deep matrix models, transformers, multimodal fusion, and foundation models for clinician-centered repurposing [18, 20]. Review papers on deep learning in drug discovery and drug repurposing acted as citation bridges by connecting molecular representation learning with biomedical network applications [10]. By 2025, deep learning papers had the highest average annual citation velocity, although many remained computationally validated rather than clinically evaluated.

*Matrix Factorization and Collaborative Filtering*

Matrix factorization and collaborative filtering represented a smaller but methodologically important cluster, accounting for approximately 8% of algorithm-focused records. These studies framed drug repurposing as a recommendation problem in which missing drug–disease or drug–target associations are inferred from low-rank structure in known interaction matrices. Low-rank matrix approximation and randomized algorithms provided one of the clearest examples of this approach in the

bibliographic network [21]. Although the cluster's publication share declined after 2020, its concepts were absorbed into later graph embedding and neural collaborative filtering models.

*Hybrid and Ensemble Approaches*

Hybrid and ensemble approaches accounted for approximately 24% of the algorithmic corpus and became the dominant methodological direction after 2022. These papers combined network topology, transcriptomic signatures, chemical similarity, drug–target interactions, biomedical text mining, and clinical phenotypes in a single prioritization pipeline. Studies such as HeTDR and structure-based AI-aided repurposing illustrate how hybrid models integrate heterogeneous evidence rather than relying on one data layer [14, 22]. The thematic map placed hybrid methods at the center of the field, linking earlier network and signature clusters to newer knowledge-graph and foundation-model communities.

**Table 1** organizes the major algorithmic communities in AI-assisted drug repurposing according to their bibliometric position, evidence logic, translational strength, and interpretive limitations.

**Table 1**. Algorithmic Communities and Bibliometric Functions in AI-Assisted Drug Repurposing

| Algorithmic community | Bibliometric position in the field | Primary evidence logic | Typical data layers | Translational strength | Main interpretive limitation | Strategic value for the field |
|---|---|---|---|---|---|---|
| Network-based inference and proximity methods | Earliest cohesive cluster; approximately 21% of algorithm-focused papers; strongly visible during COVID-19 citation bursts | Therapeutic candidates are prioritized by proximity among drugs, diseases, proteins, genes, and pathways | Protein–protein interaction networks; drug–target links; disease modules; pathway databases | Strong for mechanistic plausibility and rapid hypothesis generation | Network completeness and biological context strongly affect predictions | Provides the intellectual backbone for later graph-based and multimodal repurposing models |
| Transcriptomic signature matching | Historically influential cluster; approximately 17% of algorithmic corpus; anchored by Connectivity Map/L1000 terminology | Drugs are prioritized when perturbation signatures reverse or modify disease-associated expression patterns | Gene-expression profiles; L1000 perturbation signatures; disease transcriptional signatures; oncology datasets | Useful for mechanism-oriented candidate screening, especially in cancer and inflammatory disease contexts | Expression reversal does not always translate into therapeutic efficacy or safety | Links molecular perturbation evidence to drug prioritization and remains an important validation layer |
| Knowledge-graph embedding and inference | Rapidly expanding after 2020; approximately 18% of algorithmic studies by 2025 | Drug repurposing is framed as link prediction across heterogeneous biomedical entities | Drug–gene–disease–phenotype graphs; ontologies; literature-derived relationships; semantic triples | Strong potential for explainable candidate ranking and cross-domain evidence integration | Graph construction choices, relation quality, and leakage risks can inflate apparent performance | Bridges traditional network medicine with modern representation learning and explainable AI |
| Deep learning and representation learning | Fastest-growing community; increased to 91 papers in 2025 | Candidate drugs are prioritized through learned representations of molecules, targets, diseases, graphs, or multimodal biomedical evidence | Molecular graphs; drug–target matrices; biomedical text; omics; clinical features; graph embeddings | High predictive capacity and flexibility across data types | Many models remain retrospectively validated and poorly connected to clinical testing | Drives technical innovation and enables foundation-model-style repurposing systems |
| Matrix factorization and collaborative filtering | Smaller but methodologically important cluster; approximately 8% of algorithm-focused records | Missing drug–disease or drug–target associations are inferred from low-rank structure in known interaction matrices | Drug–disease matrices; drug–target matrices; similarity matrices; interaction histories | Useful for benchmarkable recommendation tasks | Often limited by sparse labels, historical bias, and weak mechanistic explanation | Supplies recommendation-system logic later absorbed into graph embedding and neural collaborative filtering |
| Hybrid and ensemble approaches | Central and dominant after 2022; approximately 24% of algorithmic corpus | Multiple evidence layers are combined to improve robustness and candidate prioritization | Network topology; transcriptomics; chemical similarity; drug–target data; biomedical text; clinical phenotypes | Highest translational promise because it approximates real-world evidence triangulation | Integration can obscure which evidence layer drives the prediction | Represents the field's movement from single-method prediction toward multimodal therapeutic reasoning |

*Results – Collaboration and Geographic Patterns*

*Co-Authorship Networks*

The co-authorship network contained 5,892 unique authors and 14,376 co-authorship links, with a modularity score of 0.71, indicating clearly separated but interconnected communities. The largest author communities corresponded to network medicine, deep learning, knowledge-graph informatics, and pharmacogenomic repurposing, while review authors served as intellectual bridges across methodological clusters [23, 24]. Translational papers had larger median team sizes than purely computational papers, with a median of 9 authors compared with 6, reflecting the need for clinical, biological, and informatics expertise. The densest collaboration bridges appeared around COVID-19 studies and large biomedical resources, where multi-institutional authorship was common [1, 16].

*Geographic Hotspots*

Country-level analysis identified the United States, China, the United Kingdom, Germany, India, Canada, Italy, Finland, South Korea, and Australia as the ten most productive contributors. The United States led total citations and international collaboration links, while China showed the fastest publication growth after 2020, particularly in deep learning and graph-based repurposing [20, 25]. When output was normalized by population, Finland, Singapore, Switzerland, and the Netherlands showed high relative productivity, consistent with strong bioinformatics and computational pharmacology infrastructures. Low- and middle-income countries contributed less than 11% of the total corpus despite the potential public-health value of lower-cost repurposing strategies.

*International vs. Domestic Collaboration*

Internationally co-authored papers represented 34.7% of the dataset and had a mean citation impact 1.42 times higher than domestic-only papers. Cross-border collaborations were especially common in studies involving curated resources, COVID-19 target identification, and translational rare-disease repurposing, where access to diverse datasets and domain expertise improved visibility [2, 26]. Domestic collaboration dominated in algorithm-development papers, particularly those focused on benchmarking new prediction models against existing datasets. The collaboration overlay map showed increasing internationalization after 2020, but also revealed that clinical validation networks remained less globally integrated than computational networks.

*Results – Citation and Co-Citation Networks*

*Most Cited Papers and Intellectual Base*

The most cited papers in the corpus were not necessarily those reporting clinical implementation but those defining reusable resources, conceptual frameworks, or broadly applicable AI methods. The next-generation Connectivity Map and Drug Repurposing Hub formed two of the strongest bibliographic anchors because they supplied standardized perturbational and compound-level evidence for downstream repurposing studies [1, 12]. DrugBank 5.0 also appeared as a high-frequency cited reference, particularly in knowledge-graph, drug–target interaction, and benchmark-construction papers [2]. Among conceptual works, the Nature Reviews Drug Discovery review on repurposing challenges and recommendations shaped the intellectual framing of translational barriers across the field [9].

*Co-Citation Clusters*

Co-citation analysis identified five major intellectual clusters corresponding to transcriptomic signature matching, network medicine, knowledge-graph inference, deep learning, and translational repurposing strategy. The network medicine cluster was anchored by systematic biomedical knowledge integration and COVID-19 network proximity studies, which were frequently cited together in papers using graph-based disease-module reasoning [3]. The deep learning cluster linked general drug discovery reviews with repurposing-specific representation-learning studies, showing how methodological advances in molecular AI diffused into repositioning workflows [10]. A smaller translational-readiness cluster grouped reviews and perspectives on the persistent gap between computational prediction and therapeutic adoption [9, 27].

*Citation Bursts*

Citation burst detection showed three major waves of attention during the 2017–2025 period. The first wave, from 2017 to 2019, centered on biomedical resources and early network-integration papers, including Connectivity Map, DrugBank, and heterogeneous network prediction studies [2, 4, 12]. The second wave, from 2020 to 2021, was dominated by COVID-19 repurposing, especially SARS-CoV-2 protein interaction mapping and open-data deep learning workflows [15, 16]. The third wave, from 2023 to 2025, shifted toward knowledge graphs, graph transformers, and foundation models, with HGTDR and clinician-centered foundation-model work appearing as recent burst references [18, 25].

*Results – Translational Trends from Computation to Clinic*

*Volume of Translational Papers*

Only 126 of the 1,644 documents, or 7.7%, met the operational definition of translational papers reporting clinical validation, prospective evaluation, real-world evidence, or regulatory-facing evidence. This proportion increased from 4.1% in 2017–2019 to 9.6% in 2020–2022, largely because COVID-19 created urgent demand for rapid repurposing pipelines [15].

Translationally oriented rare-disease studies also became visible, particularly where human expert review and machine intelligence were combined to prioritize candidates for conditions with limited therapeutic options [26]. Nevertheless, the bibliometric gap between computational output and clinical validation remained substantial across all algorithmic clusters.

*Typical Pathways to Translation*

The most common translational pathway was computational prioritization followed by retrospective biological or clinical plausibility assessment, rather than prospective randomized evaluation. A smaller group of studies connected AI prediction to observational data, electronic health records, or clinical expert panels, reflecting the emerging role of real-world validation in repurposing pipelines [18, 26]. Structure-based and integrated in silico approaches were often positioned as preclinical triage tools that could reduce the number of compounds requiring experimental testing [14, 28]. Across the translational sub-corpus, only 18 papers reported evidence directly linked to trial initiation or trial-adjacent evaluation.

*Translational Success Stories*

The most visible translational footprint was associated with COVID-19, where repurposing predictions and mechanistic maps rapidly influenced clinical and experimental prioritization. SARS-CoV-2 protein interaction mapping produced a highly cited bridge between molecular virology and drug screening, creating a dense co-citation cluster with network medicine and open-data deep learning papers [15, 16]. Baricitinib-related repurposing discussions contributed to the broader pandemic-era citation burst, although much of the bibliometric signal came from computational and mechanistic studies rather than definitive clinical-trial papers. These examples show that AI-assisted repurposing can generate translational momentum, but sustained validation depends on coordinated clinical testing infrastructure.

*Results – Thematic Evolution and Emerging Topics*

*Temporal Keyword Dynamics*

Temporal keyword analysis showed a clear shift from "network," "drug–target interaction," and "gene expression" in 2017–2019 toward "knowledge graph," "graph neural network," "transformer," and "foundation model" in 2023–2025. The transcriptomic cluster remained anchored to Connectivity Map and L1000 terminology, but its relative prominence declined as graph-based biomedical reasoning became more central [12, 17]. Network-based terms persisted across the entire period because many newer models still rely on graph-structured evidence, even when framed as deep learning or representation learning [4, 20]. The overlay visualization therefore suggested methodological convergence rather than simple replacement of older approaches.

*Emerging Research Fronts*

The most active emerging fronts were multimodal foundation models, explainable repurposing, graph transformers, and integration of computational predictions with real-world clinical evidence. Heterogeneous graph transformers represented a visible technical frontier because they extended earlier graph embedding approaches into attention-based architectures suitable for large biomedical networks [25]. Clinician-centered foundation models marked a second frontier by explicitly designing AI outputs around human decision-making and translational interpretability [18]. Reviews of machine learning applications in drug discovery indicate that these trends are part of a broader movement from task-specific prediction toward integrated, decision-support-oriented AI systems. **Figure 3** synthesizes publication growth, algorithmic communities, citation infrastructure, collaboration patterns, and the translational bottleneck shaping AI-assisted drug repurposing from 2017 to 2025.

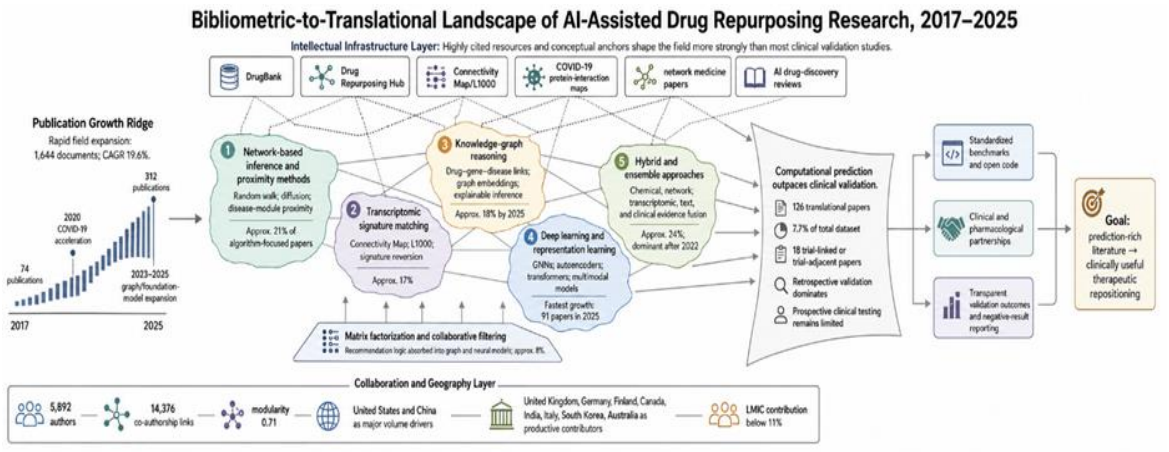


**Figure 3.** Bibliometric-to-Translational Landscape of AI-Assisted Drug Repurposing Research

*Maturing Field with Converging Methodologies*

The bibliometric structure indicates that AI-assisted drug repurposing is maturing from isolated algorithm families into a convergent, multimodal field. Early distinctions between network proximity, transcriptomic signature matching, and matrix

completion are becoming less rigid as hybrid methods combine graph topology, perturbational profiles, chemical features, and clinical evidence [14, 22]. Deep learning studies now frequently absorb concepts from recommender systems, biomedical networks, and knowledge graphs rather than replacing them outright [17, 20]. This convergence explains why hybrid and ensemble approaches occupied the center of the thematic map in the 2023–2025 period.

*The Dominance of Method Papers Over Application Papers*

The field remains dominated by method-development papers, benchmark demonstrations, and retrospective validation studies. Highly cited methodological work such as deepDR and iDrug illustrates the productivity of computational innovation, but it also shows that many influential papers stop at in silico or database-level validation [19, 20]. Reviews of drug repurposing and AI-assisted discovery repeatedly emphasize that predictive performance on retrospective benchmarks does not guarantee clinical usefulness [9, 27]. The bibliometric imbalance is therefore structural: papers proposing algorithms are easier to publish and cite than papers carrying predictions through costly biological and clinical validation.

*The Translational Gap as a Structural Weakness*

The translational gap appears to arise from weak integration between computational authorship networks and clinical testing communities. Co-authorship patterns showed that translational papers had larger teams and more cross-disciplinary links, consistent with the need for clinicians, pharmacologists, trialists, and informaticians to collaborate [18, 26]. Yet most algorithmic papers were produced within computationally cohesive groups with limited evidence of downstream clinical partnerships. This disconnect helps explain why the corpus contains many high-performing prediction models but comparatively few studies reporting prospective validation or regulatory engagement.

*Geographic Disparities in Research Output*

Geographic analysis showed strong concentration of output in high-income countries, especially the United States, China, the United Kingdom, Germany, Canada, and several smaller high-productivity European countries. This concentration partly reflects access to large biomedical databases, computing infrastructure, and interdisciplinary funding streams, which are central to knowledge-graph and deep-learning repurposing research [17, 25]. However, the low contribution from low- and middle-income countries is notable because drug repurposing could be especially valuable where new drug development is financially inaccessible. More equitable global participation would broaden disease priorities beyond oncology, COVID-19, and well-funded rare-disease programs.

*The Role of COVID-19 as an Accelerant*

COVID-19 acted as a major accelerant, producing a temporary surge in publications, citations, and international collaborations. The pandemic elevated network medicine, protein interaction mapping, and open-data deep learning as rapid-response tools for therapeutic prioritization [15, 16]. It also exposed limitations of fast computational repurposing, including the risk that large numbers of predictions may outpace experimental verification and clinical trial capacity. Bibliometrically, COVID-19 increased the visibility of AI-assisted repurposing while also making the translational gap more apparent.

*Implications for Funding and Policy*

Funding and policy strategies should treat AI-assisted repurposing as a translational pipeline rather than only a computational discovery activity. Dedicated support is needed for prospective validation, interoperable benchmarks, negative-result reporting, and clinically meaningful outcome tracking, because these elements are underrepresented in the citation network [9, 27]. Public investment in shared infrastructures such as curated compound libraries and perturbational databases can also generate broad downstream value, as shown by the citation centrality of DrugBank, the Drug Repurposing Hub, and Connectivity Map [1, 2, 12]. Policy incentives should therefore reward validated therapeutic impact as well as algorithmic novelty.

**Table 2** translates the bibliometric findings into a readiness matrix that specifies the evidence upgrades needed to move AI-assisted drug repurposing from prediction-rich research toward clinically actionable therapeutic repositioning.

**Table 2.** Translational Readiness Matrix for AI-Assisted Drug Repurposing Research

| Translational dimension | Bibliometric signal observed in the review | Current weakness | Required evidence upgrade | Recommended reporting or policy action | Expected impact on field maturity |
|---|---|---|---|---|---|
| Computational prediction | Large and rapidly expanding method-development literature; strong growth in graph, deep learning, and hybrid models | Benchmark performance is often treated as evidence of usefulness | External validation across independent datasets, disease areas, and drug classes | Require transparent dataset provenance, code availability, negative controls, and leakage checks | Improves reproducibility and reduces overclaiming of model performance |
| Biological plausibility | Network medicine, transcriptomic signatures, and pathway- | Mechanistic plausibility is often inferred rather | Orthogonal biological validation using perturbation assays, | Ask authors to distinguish computational plausibility from | Strengthens confidence in candidate |

| | | | | | |
|---|---|---|---|---|---|
| | based reasoning are highly visible in citation clusters | than experimentally demonstrated | pathway readouts, or target-engagement evidence | experimentally supported mechanism | prioritization before clinical investment |
| Clinical validation | Only 126 of 1,644 papers, or 7.7%, met the translational-paper definition | Clinical evidence remains a small fraction of the publication record | Retrospective EHR validation, observational studies, pragmatic trials, or trial-adjacent evaluation | Use structured article labels for prediction-only, preclinical validation, real-world evidence, and prospective testing | Makes the computation-to-clinic gap measurable and easier to monitor |
| Trial linkage | Only 18 papers reported evidence directly linked to trial initiation or trial-adjacent evaluation | Few computational predictions progress to formal clinical testing | Predefined candidate-selection criteria and documented handoff to clinical or regulatory teams | Encourage publication of candidate triage protocols and trial-readiness criteria | Connects algorithmic output to therapeutic development decisions |
| Collaboration structure | Translational papers had larger median teams than computational papers | Algorithmic groups are often weakly connected to clinicians, pharmacologists, and trialists | Integrated teams including informaticians, disease specialists, pharmacologists, and clinical investigators | Funding calls should require translational partnership plans and validation milestones | Increases the probability that predictions move beyond retrospective analysis |
| Geographic equity | High-income countries dominate output; LMIC contribution remains below 11% | Disease priorities may reflect data-rich and well-funded settings | Cross-border datasets, neglected-disease partnerships, and resource-sensitive validation models | Create international repurposing networks focused on public-health priorities and affordable therapies | Broadens the relevance of AI-assisted repurposing beyond dominant research economies |
| Bibliometric traceability | Citation networks privilege resources, reviews, and methods over validation studies | Citation impact does not equal clinical impact | Combined assessment of citations, validation outcomes, trial progression, and patient-relevant endpoints | Journals and funders should track validation status alongside citation metrics | Shifts incentives from algorithmic novelty toward demonstrated translational value |
| Negative-result reporting | Positive computational predictions dominate the visible literature | Failed validation and non-replicated candidates are underreported | Public reporting of unsuccessful candidates, invalidated predictions, and benchmark failures | Establish negative-result repositories or journal sections for repurposing validation | Reduces publication bias and improves future model calibration |

*Comparison with Other Bibliometric Studies*

The findings align with prior bibliometric analyses showing rapid growth of AI in drug discovery and increasing centrality of machine learning, deep learning, and data integration. Earlier bibliometric studies identified rising publication volume and strong concentration in a small set of productive countries and journals, patterns that were also observed in this repurposing-focused dataset [5–8]. The present analysis differs by emphasizing translational trajectories and by separating algorithmic clusters into network-based, signature-based, knowledge-graph, deep-learning, matrix-factorization, and hybrid communities. It also shows that bibliometric visibility is not equivalent to clinical readiness, a distinction that is less prominent in broader AI-drug-discovery mapping studies.

*Limitations*

*Database and Coverage Limitations*

This review relied on PubMed, Scopus, and Web of Science, which provide strong coverage of biomedical and computational journals but may miss grey literature, preprints, regulatory documents, trial registry updates, and non-English clinical reports. Because AI-assisted drug repurposing often moves through conference papers, software repositories, and institutional collaborations before journal publication, some early methodological signals may be undercounted [6]. The search strategy also emphasized drug repurposing and repositioning terminology, which may exclude adjacent work described as indication expansion, therapeutic switching, or computational pharmacology. These coverage limitations mean that the dataset should be interpreted as a peer-reviewed publication map rather than a complete innovation map.

*Bibliometric Indicators May Not Fully Capture Quality or Impact*

Citation counts, h-index values, and burst metrics are useful for identifying influence but do not directly measure methodological rigor, clinical validity, or patient benefit. Resource papers and broad reviews naturally accumulate more citations than narrowly focused validation studies, which can inflate their apparent dominance in co-citation maps [2, 9]. Conversely, clinically important translational work may receive fewer citations if it appears in specialist journals, trial reports, or disease-specific venues. Bibliometric indicators should therefore be combined with expert review, validation tracking, and outcome-based assessment when evaluating the real-world value of AI-assisted repurposing.

*Implications and Recommendations*

*For Researchers*

Researchers should design AI-assisted repurposing studies with validation and reproducibility in mind from the outset. Shared benchmarks, open code, transparent negative controls, and external validation datasets would make algorithmic comparisons more meaningful across network, signature, knowledge-graph, and deep-learning approaches [10, 19]. Multidisciplinary authorship should be prioritized, particularly by integrating clinicians, disease specialists, pharmacologists, and trial methodologists before candidate prioritization is finalized [18, 26]. The field would also benefit from clearer reporting of whether predictions are intended for hypothesis generation, preclinical testing, observational validation, or clinical trial prioritization.

*For Funders and Institutions*

Funders and institutions should create targeted programs that bridge the space between computational prediction and clinical evaluation. Bibliometric evidence suggests that the field has sufficient algorithmic productivity but insufficient support for prospective validation, translational partnerships, and cross-border clinical networks [9, 27]. Funding calls could require integrated teams, predefined validation milestones, and use of shared resources such as DrugBank, Connectivity Map, and curated repurposing libraries [1, 2, 12]. International partnership mechanisms would be especially valuable for extending AI-assisted repurposing toward neglected diseases and health systems with fewer new-drug development resources.

*For Journal Editors*

Journal editors can improve the bibliometric traceability and translational value of the field by encouraging standardized reporting of datasets, algorithms, validation designs, and clinical relevance. Manuscripts should clearly distinguish retrospective benchmark performance from experimental validation, real-world evidence, and prospective clinical testing, because these categories are often blurred in the literature [14, 28]. Dedicated article tags or structured abstracts for "repurposing prediction," "preclinical validation," and "clinical translation" would make future bibliometric surveillance more accurate. Journals could also encourage publication of negative validation studies, reducing citation bias toward successful computational predictions.

## Conclusion

AI-assisted drug repurposing has become a vibrant and rapidly growing research area, with five distinct algorithmic clusters driving innovation. Network-based inference, transcriptomic signature matching, knowledge-graph reasoning, deep learning, and hybrid multimodal approaches together define the field's methodological architecture.

Despite the volume and sophistication of computational work, translational outputs remain a small fraction of the literature. Most highly cited papers describe methods, resources, or conceptual frameworks rather than prospective clinical validation.

The field's structure reveals a need to strengthen bridges between computational scientists and clinical researchers. Evaluation standards should reward reproducibility, external validation, transparent reporting, and measurable real-world impact.

Coordinated bibliometric surveillance can help track whether the field is moving from prediction-rich research toward clinically useful therapeutic repositioning. Such monitoring can guide funding, collaboration, and journal policy toward work that improves both scientific quality and translational productivity.

**Acknowledgments:** None

**Conflict of interest:** None

**Financial support:** None

**Ethics statement:** None

## References

1. Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, et al. The Drug Repurposing Hub: a next-generation drug library and information resource. Nat Med. 2017;23(4):405-8.
2. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074-82.
3. Himmelstein DS, Lizee A, Hessler C, Brueggeman L, Chen SL, Hadley D, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. eLife. 2017;6:e26726.
4. Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W, et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nat Commun. 2017;8:573.
5. Sun G, Dong D, Dong Z, Zhang Q, Fang H, Wang C, et al. Drug repositioning: a bibliometric analysis. Front Pharmacol. 2022;13:974849.
6. Karger E, Kureljusic M. Using artificial intelligence for drug discovery: A bibliometric study and future research agenda. Pharmaceuticals (Basel). 2022;15(12):1492.

7. Mak KK, Wong YH, Pichika MR. Artificial intelligence in drug discovery and development. In: Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays. 2024. p. 1461-98.
8. Koçak M, Akçalı Z. The published role of artificial intelligence in drug discovery and development: A bibliometric and social network analysis from 1990 to 2023. J Cheminform. 2025;17(1):71.
9. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41-58.
10. Pan X, Lin X, Cao D, Zeng X, Yu PS, He L, et al. Deep learning for drug repurposing: Methods, databases, and applications. Wiley Interdiscip Rev Comput Mol Sci. 2022;12(4):e1597.
11. Trejo-Castro AI, Martinez-Ledesma E, Martinez-Torteya A. A bibliometric review on in silico drug repurposing: Performance analysis, science mapping and text mining (2000-2023). Heliyon. 2025;11(10).
12. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell. 2017;171(6):1437-52.
13. Sam E, Athri P. Web-based drug repurposing tools: a survey. Brief Bioinform. 2019;20(1):299-316.
14. Choudhury C, Murugan NA, Priyakumar UD. Structure-based drug repurposing: Traditional and advanced AI/ML-aided methods. Drug Discov Today. 2022;27(7):1847-61.
15. Zeng X, Song X, Ma T, Zhou J, Yu PS, He L, et al. Repurpose open data to discover therapeutics for COVID-19 using deep learning. J Proteome Res. 2020;19(11):4624-36.
16. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020;583(7816):459-68.
17. Bang D, Lim S, Lee S, Kim S. Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. Nat Commun. 2023;14:3570.
18. Huang K, Chandak P, Wang Q, Havaldar S, Vaid A, Leskovec J, et al. A foundation model for clinician-centered drug repurposing. Nat Med. 2024;30(12):3601-13.
19. Chen H, Cheng F, Li J. iDrug: Integration of drug repositioning and drug-target prediction via cross-network embedding. PLoS Comput Biol. 2020;16(7):e1008040.
20. Zeng X, Zhu S, Liu X, Zhou Y, Nussinov R, Cheng F. deepDR: a network-based deep learning approach to in silico drug repositioning. Bioinformatics. 2019;35(24):5191-8.
21. Luo H, Li M, Wang S, Liu Q, Li Y, Wang J. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. Bioinformatics. 2018;34(11):1904-12.
22. Jin S, Niu Z, Jiang C, Huang W, Xia F, Jin X, et al. HeTDR: Drug repositioning based on heterogeneous networks and text mining. Patterns (N Y). 2021;2(8).
23. Tanoli Z, Vähä-Koskela M, Aittokallio T. Artificial intelligence, machine learning, and drug repurposing in cancer. Expert Opin Drug Discov. 2021;16(9):977-89.
24. Koromina M, Pandi MT, Patrinos GP. Rethinking drug repositioning and development with artificial intelligence, machine learning, and omics. OMICS. 2019;23(11):539-48.
25. Gharizadeh A, Abbasi K, Ghareyazi A, Mofrad MRK, Rabiee HR. HGTDR: Advancing drug repurposing with heterogeneous graph transformers. Bioinformatics. 2024;40(7):btae349.
26. Challa AP, Zaleski NM, Jerome RN, Lavieri RR, Shirey-Rice JK, Barnado A, et al. Human and machine intelligence together drive drug repurposing in rare diseases. Front Genet. 2021;12:707836.
27. Anokian E, Bernett J, Freeman A, List M, Santamaría LP, Tanoli Z, et al. Machine learning and artificial intelligence in drug repurposing—challenges and perspectives. Drug Repurposing. 2024;1(1).
28. March-Vila E, Pinzi L, Sturm N, Tinivella A, Engkvist O, Chen H, et al. On the integration of in silico drug design methods for drug repurposing. Front Pharmacol. 2017;8:272508.