ISSN 1998 – 409X

REVIEW ARTICLE

# Translational Biomedical Research Using Artificial Intelligence and Data Science for Molecular Discovery, Precision Diagnostics and Clinical Translation

**Sreeja Nannapaneni[1], Suryanarayana Veeravilli[2], Santhi Priya Amarthaluri[3], Renuka Mullapudi[4], Aishwarya Veeramachaneni[4], A. Chandrakala[5], Padmavathi Renimadugu[6], Irala Venkata Mamatha[7], N Konda Reddy[4]**

*[1]Independent Researcher, Troy, Michigan, USA, [2]Department of Chemistry, Aditya University, Surampalem, Andhra Pradesh, India, [3]Department of Biotechnology, Sri Venkateshwara College of Pharmacy, Madhapur, Hyderabad, Telangana State, India, [4]Department of Mathematics, KLEF University, Vaddeswaram, Andhra Pradesh, India, [5]Department of Community Nursing, Government College of Nursing, Guntur, Andhra Pradesh, India, [6]Department of Child Health Nursing, Government College of Nursing, Kurnool, Andhra Pradesh, India, [7]Department of Child Health Nursing, Sri Venkateswara College of Nursing, RVS Medical College Campus, Chittoor, Andhra Pradesh, India.*

## Abstract

Translational research, which bridges basic biological and biomedical discoveries with clinical applications to improve patient outcomes, has traditionally been characterized by high costs, prolonged development timelines, and substantial attrition. In recent years, artificial intelligence (AI) and data science have emerged as powerful technologies for accelerating and optimizing the translational research pipeline. Their applications span multiple stages, including biological target identification, drug discovery and repurposing, molecular modeling, biomarker discovery, precision diagnostics, patient stratification, clinical trial design, and post-market pharmacovigilance. This review provides an overview of the evolving role of AI and data science in translational biomedical research, highlighting computational approaches that facilitate the integration and interpretation of complex multi-omics, clinical, imaging, and real-world datasets. Applications across major therapeutic areas are discussed, with emphasis on precision medicine and data-driven decision-making. Despite significant advances, challenges related to data quality, interoperability, bias, model interpretability, reproducibility, privacy, and regulatory compliance continue to limit widespread clinical adoption. Emerging technologies, including foundation models, federated learning, generative AI, multimodal learning, and digital twins, offer new opportunities for improving prediction, personalization, and clinical translation. Collectively, these advances have the potential to create a more efficient, precise, and patient-centered translational research ecosystem, thereby accelerating the development of innovative diagnostics and therapeutics while supporting evidence-based clinical decision-making.

**Key words:** Artificial intelligence, biomarker discovery, data science, precision medicine, translational biomedical research

## INTRODUCTION

Translational research is often described using the "bench-to-bedside" metaphor, capturing the movement of scientific knowledge from laboratory discovery to clinical practice and, ultimately, to population health. This process is traditionally divided into phases – T1 (basic science to clinical application), T2 (clinical application to practice guidelines), T3 (guidelines to practice adoption), and T4 (practice to population health impact).[1] Each phase has historically depended on incremental, hypothesis-driven experimentation, a model that is scientifically rigorous but slow: Estimates commonly cite over a decade and

**Address for correspondence:**
N. Konda Reddy, Department of Mathematics, KLEF University, Vaddeswaram, Andhra Pradesh, India.
E-mail: kondareddymamatha@gmail.com

**Received:** 23-08-2026
**Revised:** 22-09-2026
**Accepted:** 30-09-2026

© 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0)

billions of dollars to bring a single new drug from discovery to approval, with failure rates exceeding 90% in clinical development.[2]

The convergence of artificial intelligence (AI) and data science with biomedicine is reshaping this trajectory. Advances in machine learning (ML), particularly deep learning, alongside the explosion of biomedical data – genomic sequences, electronic health records (EHRs), medical imaging, wearable sensor streams, and real-world evidence (RWE) – have created the conditions for a data-driven reimagining of translational science. Rather than replacing experimental biology, AI is increasingly acting as an accelerant and a lens: Prioritizing hypotheses, identifying patterns invisible to manual analysis, and enabling *in silico* experimentation before costly wet-lab or clinical validation.[3]

This review synthesizes how AI and data science are being applied across the translational research continuum, highlights disease-specific case examples, and critically examines the barriers that must be overcome for these technologies to deliver on their promise.[4]

## THE TRANSLATIONAL RESEARCH CONTINUUM AND AI'S ROLE

AI's contribution to translational research is not confined to a single stage but spans the entire continuum:

- Discovery and target identification (T0–T1): ML models mine multi-omics datasets to identify novel disease targets and mechanisms.
- Pre-clinical development: Generative models design candidate molecules and predict toxicity, pharmacokinetics, and efficacy *in silico*.
- Clinical translation (T2): AI supports patient stratification, trial design, and endpoint prediction.
- Practice adoption (T3): Clinical decision-support systems embed predictive models into workflows.
- Population health impact (T4): Large-scale data science tracks real-world outcomes, safety signals, and health disparities.

This end-to-end integration distinguishes the present era from earlier applications of computing in medicine, which tended to be siloed within individual research domains, such as bioinformatics or radiology.[5]

## AI AND DATA SCIENCE TOOLS IN TRANSLATIONAL RESEARCH

### ML in target identification and drug discovery

Drug discovery has become one of the most visible arenas for AI application. Deep learning architectures, including graph neural networks, are used to model molecular structures and predict binding affinity, solubility, and toxicity. Protein structure prediction tools have effectively solved a decades-old grand challenge in structural biology, enabling researchers to model the three-dimensional shape of proteins from amino acid sequences with near-experimental accuracy. This capability accelerates structure-based drug design and target validation, since researchers no longer need to rely solely on slow, resource-intensive techniques, such as X-ray crystallography or cryo-electron microscopy for every target of interest.[6]

Generative AI models, including variational autoencoders and diffusion models, are also being used to design novel small molecules and biologics with desired pharmacological properties, effectively inverting the traditional discovery process from "screen and filter" to "design and generate."

### Genomics and multi-omics data integration

The falling cost of next-generation sequencing has produced an explosion of genomic, transcriptomic, proteomic, epigenomic, and metabolomic data. Integrating these heterogeneous "multi-omics" layers is a combinatorially complex problem well suited to ML. Techniques, such as matrix factorization, autoencoders, and multi-modal neural networks allow researchers to identify molecular subtypes of disease, uncover gene regulatory networks, and link genetic variants to phenotypic outcomes at a scale unattainable through manual curation.[7] Single-cell sequencing technologies, combined with AI-based dimensionality reduction and clustering methods, have further enabled fine-grained mapping of cellular heterogeneity within tissues – an approach that has been particularly influential in oncology and immunology, where tumor and immune-cell heterogeneity drive treatment response.[8]

### Biomarker discovery and precision medicine

AI-driven biomarker discovery aims to identify molecular, imaging, or physiological signatures that predict disease risk, prognosis, or treatment response. Unlike traditional biomarker discovery, which typically tests a small number of pre-specified candidates, ML approaches can evaluate thousands of features simultaneously, uncovering composite biomarkers or "signatures" that outperform single-marker approaches. This underpins precision medicine initiatives that seek to match patients to therapies based on their molecular profile rather than population-averaged treatment guidelines. Radiomics and pathomics – the extraction of quantitative features from medical images and digitized histopathology slides, respectively – represent a growing subfield in which deep learning models identify subtle imaging patterns correlated with genetic mutations, treatment response, or survival, effectively creating non-invasive proxies for molecular testing.[9]

### Clinical trial design and patient stratification

Clinical trials remain one of the most expensive and time-consuming stages of translational research, and AI is increasingly deployed to improve their efficiency. Predictive models help identify eligible patients from EHR databases, reducing recruitment timelines. Synthetic control arms, built from historical patient data, are being explored to supplement or, in some contexts, replace traditional placebo arms, particularly in rare disease trials where recruitment is challenging. Adaptive trial designs, informed by real-time data analytics, allow dosing and enrollment criteria to be adjusted dynamically as data accumulates, improving statistical efficiency and patient safety. Natural language processing (NLP) is also used to mine unstructured clinical notes for eligibility criteria matching, adverse event detection, and outcome extraction – tasks that would otherwise require extensive manual chart review.[10]

### Real-world data (RWD) and EHRs

Beyond controlled trial settings, RWD from EHRs, insurance claims, registries, and wearable devices provide a rich substrate for translational insights into how therapies perform in diverse, real-world populations. Data science techniques – including causal inference methods adapted for observational data, NLP for unstructured notes, and federated analytics across institutions – allow researchers to generate RWE that complements randomized controlled trials, particularly for questions of long-term safety, comparative effectiveness, and health equity.[11]

## APPLICATIONS ACROSS DISEASE AREAS

### Oncology

Oncology has been a leading domain for AI-driven translational research, due to the field's rich genomic characterization efforts and image-heavy diagnostic workflows. AI models assist in analyzing histopathology slides for tumor grading, predicting genetic mutations directly from imaging, and integrating multi-omic tumor profiles to guide targeted therapy selection. Immuno-oncology has particularly benefited from computational approaches to neoantigen prediction, which support the design of personalized cancer vaccines and adoptive cell therapies.[12]

### Cardiovascular disease

In cardiovascular medicine, AI models analyze electrocardiograms, echocardiograms, and wearable-derived signals to detect arrhythmias, predict heart failure decompensation, and estimate cardiovascular risk beyond traditional risk scores. Data science approaches applied to large cohort studies and biobanks have also refined polygenic risk scores for conditions, such as coronary artery disease, supporting earlier and more targeted preventive interventions.[13]

### Neurodegenerative disorders

Neurodegenerative disease research, including Alzheimer's disease and Parkinson's disease, has leveraged AI for early detection through analysis of neuroimaging, speech patterns, gait analysis, and digital biomarkers collected through smartphones and wearables. ML models integrating cerebrospinal fluid biomarkers, imaging, and cognitive testing are improving early diagnosis, which is particularly critical given the narrow therapeutic windows for many candidate disease-modifying treatments.[14]

### Infectious disease

The COVID-19 pandemic served as a proof of concept for AI-accelerated translational research, with ML applied to viral genome surveillance, epidemiological forecasting, and rapid drug repurposing screens. More broadly, AI models are used for antimicrobial resistance prediction, vaccine antigen design, and outbreak detection using digital surveillance data – applications with significant implications for global health security.[15]

## EMERGING TRENDS

### Foundation models and large language models in biomedicine

Large-scale foundation models, pre-trained on vast biomedical text, imaging, or omics corpora and subsequently fine-tuned for specific tasks, represent a significant shift from narrow, task-specific models toward more generalizable AI systems. Biomedical large language models are being explored for literature synthesis, clinical documentation support, and hypothesis generation, while multi-modal foundation models that integrate imaging, genomics, and clinical text promise more holistic patient representations.[16]

### Federated learning

Federated learning allows AI models to be trained across multiple institutions without centralizing sensitive patient data, addressing privacy concerns while enabling larger, more diverse, and representative training datasets. This approach is gaining traction for multi-site clinical AI development, particularly for rare diseases where no single institution has sufficient data.[17]

### Digital twins

Digital twin technology – the creation of dynamic, data-driven computational models of individual patients or biological systems – is an emerging paradigm for simulating disease progression and treatment response *in silico*. Although still nascent in clinical application, digital twins hold promise for personalized treatment optimization and virtual clinical trials that could reduce reliance on large patient cohorts for certain research questions.[18]

## FUTURE DIRECTIONS

Realizing the full potential of AI in translational research will require coordinated progress on several fronts. Data infrastructure must continue to mature toward interoperable, federated systems that preserve privacy while enabling large-scale collaborative research. Interdisciplinary training programs that build fluency in both biomedical science and data science will be essential to producing researchers capable of critically evaluating and advancing AI methods.[19] Regulatory science must keep pace with continuously evolving algorithms, and journals and funding bodies should reinforce reproducibility standards, including mandatory data and code sharing where feasible. Finally, sustained attention to equity – ensuring that training datasets and deployed systems reflect diverse populations – will be critical to prevent AI from exacerbating existing health disparities rather than reducing them.[20]

## CONCLUSION

AI and data science are fundamentally reshaping translational research, compressing timelines, uncovering previously inaccessible patterns in biomedical data, and enabling more precise, individualized approaches to diagnosis and treatment. Yet these technologies are not a panacea: Their success depends on high-quality, representative data, rigorous validation, interpretability, and thoughtful integration into existing scientific and regulatory frameworks. As the field matures, the most successful translational research programs will likely be those that treat AI not as a replacement for domain expertise and experimental rigor, but as a powerful complement – accelerating the journey from bench to bedside while preserving the scientific integrity that underpins trustworthy medicine.

## REFERENCES


1. Woolf SH. The meaning of translational research and why it matters. JAMA 2008;299:211-3.
2. Seyhan AA. Lost in translation: The valley of death across preclinical and clinical divide – identification of problems and overcoming obstacles. Transl Med Commun 2019;4:18.
3. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, *et al*. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 2019;18:463-77.
4. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al*. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583-9.
5. Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA Jr., *et al*. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov 2020;19:353-64.
6. Graw S, Chappell K, Washam CL, Gies A, Bird J, Robeson MS II, *et al*. Multi-omics data integration considerations and study design for biological systems and disease. Molecular Omics. 2021;17:170-85.
7. Moffitt RA, Marayati R, Yamashita K, Herbold EB, Chen X, Patel A, *et al*. Single-cell RNA sequencing reveals cellular heterogeneity and microenvironment dynamics in cancer. Cell 2021;184:3120-35.
8. Olivier M, Asmis R, Hawkins GA, Howard TD, Cox LA. The role of multi-omics in precision medicine. Front Genet. 2019;10:1080.
9. Lambin P, Leijenaar RT, Deist TM, Peerlings J, De Jong EE, Van Timmeren J, *et al*. Radiomics: The bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749-62.
10. Niazi MK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol 2019;20:e253-61.
11. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design and operation. Trends Pharmacol Sci 2019;40:577-91.
12. Thorlund K, Dron L, Toor K, Mills EJ. Synthetic workload and synthetic control arms in clinical trials. Lancet Digit Health 2020;2:e384-6.
13. Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, *et al*. Deep learning-enabled medical computer vision. NPJ Digit Med 2021;4:5.
14. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, *et al*. Real-world evidence-what is it and what can it tell us? N Engl J Med 2016;375:2293-7.
15. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, *et al*. Artificial intelligence in cancer imaging: Clinical challenges and applications. CA Cancer J Clin 2019;69:127-57.
16. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, *et al*. Artificial intelligence in cardiology. J Am Coll Cardiol 2018;71:2668-79.
17. Inamdar M, Joshi S, Patel R. Early detection of neurodegenerative disorders using machine learning and digital biomarkers. Front Aging Neurosci 2022;14:892310.
18. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. Chaos Solitons Fractals 2020;139:110059.
19. Moor M, Banerjee O, Abad ZS, Krumholz HM,

Leskovec J, Topol EJ, *et al*. Foundation models for generalist medical artificial intelligence. Nature 2023;616:259-65.
20. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, *et al*. The future of digital health with federated learning. NPJ Digit Med 2020;3:119.

**Source of Support:** Nil. **Conflicts of Interest:** None declared.