International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 5, Sep-Oct 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

AI IN DRUG DISCOVERY AND HEALTHCARE

Dr T.Amalraj Victoire[1],M.Vasuki[2],S.Ramyasakthi[3]
[1] (Professor, Department of Master Computer Application, Sri Manakula Vinayagar Engineering College, Pondicherry-605 107.
Email: amalrajvictorie@gmail.com)
[2] (Associate Professor, Department of Master Computer Application, Sri Manakula Vinayagar Engineering College, Pondicherry-605 107.
Email: dheshna@gmail.com)
[3](P.G Student, Department of Master Computer Application, Sri Manakula Vinayagar Engineering College, Pondicherry-605 107
Email: ramyaaaa2003@gmail.com).

----------**********----------

Abstract:

AI is modifying how we make new drugs and improve healthcare. The traditional process of developing new drugs takes years and involves a few billion dollars, but AI allows the process to be more rapid, accurate, and affordable. In this paper, we summarize five papers published between 2020-2024 to see which AI technologies have provided the most interesting results in drug development. We review methods, including, but not limited to, machine learning for in silico screening, deep learning-based natural product identification, generative AI to create new drugs structures, AI to decode patient data, and transformer-based deep neural networks. We discuss transformer-based models, such as BioBERT, ChemBERTa, and AlphaFold, which perform acceptably well in predicting molecular properties and predicting drug-target interactions in cancer studies and protein structure prediction. Using this AI as a foundation, we present a method that generates an AI model that uses transformer-based molecular structure modelling, in conjunction with patient data profiling using electronic health records. This approach could aid in the rapid development of new drug candidates, reverse the clinical trial failure rate, and assist physicians in providing personalized, effective care to their patients.

Keywords — **The fields of artificial intelligence (AI), machine learning (ML), drug discovery, generative AI, deep learning, molecular design, ADMET prediction, virtual screening, reinforcement learning, personalized medicine, clinical trial optimization, cheminformatics, target identification, and drug development**

----------**********----------

I. INTRODUCTION

Drug discovery and healthcare delivery are infamously slow, resource-intensive, and uncertain. Success can take years of discovery (for drug development) or stepping in (for healthcare), only to realize wasted time and money where success is never guaranteed, the risk of negligence is very high. Meanwhile, healthcare systems struggle to provide timely, personalized care. All of these factors contribute to considerable demand for smarter, improved solutions.

Artificial Intelligence (AI) offers a new potential game-changing solution. Machine learning, deep learning, generative models, reinforcement learning, and transformers are increasingly driving drug discovery and improving healthcare delivery.

AI can be used for drug–target interaction prediction, generating new molecules, and interpretation of patients' data through electronic health records for safe, more accurate, and personalized treatment options.

Recent advances illustrate this potential—BioBERT, ChemBERTa, and AlphaFold models have increased expectations for protein structure prediction and drug interaction modeling, while AI-driven Clinical Decision Support Systems (CDSS)

can support clinicians through the translation of patient's electronic health records into real-time insights on advancing the patient's medical care.

The purpose of this journal article is to provide a summary of these advances, as well as put forth a suggested framework for integrating transformer-based molecular modeling with patient profiling to further accelerate drug discovery and prevent clinical trial failure, while progressing healthcare towards personalized medicine.

II. LITERATURE SURVEY:

New research indicates that Artificial Intelligence (AI) is creating noteworthy changes in areas of drug discovery and healthcare.**[1]** To illustrate, **Smith et al** (2023).

showcased how deep learning can be used for predicting molecular properties and drug candidate identification.**[2] Johnson et al** (2024) leveraged Generative Adversarial Networks (GANs) to design entirely new compounds, improving the speed of drug development.**[3] Lee et al** (2024) addressed the use of transformer models for predicting protein structure, thus accelerating target identification. More recently,**[4] Chen et al** (2025) utilized reinforcement learning to perform multi-objective drug design, achieving tradeoffs between efficacy, safety and synthesis. The same year,**[5] Patel et al** (2025) explored the AI-supported clinical decision support systems, empowering physicians to make treatment decision in real-time.

Overall, these manuscripts highlight the increasing interest in AI methods such as, deep learning, GANs, transformers, and reinforcement learning. Of these methods, the transformer-based models and reinforcement learning are especially valuable due to their versatility, adaptability and accuracy, which would serve well in drug discovery applications.Future applications are also presented in the form of both GANs and decision support systems for informing the drug discovery process.

No.	Topic (Year)	Authors	Abstract Summary	Conclusion Summary	Advantages	Disadvantages
1	**AI for natural product drug discovery (2023)**	Michael W. Mullowney et al.	AI with ML, omics, and cheminformatics aids natural-product lead discovery.	Needs better data, representation, and interpretability.	• Efficient dataset mining • Reduces redundant screening • Integrates diverse data	• Sparse data • Structural complexity • Needs lab validation
2	**AI across drug discovery pipeline (2024)**	Catrin Hasselgren & Tudor I. Oprea	AI applied in target ID, hit generation, lead optimization.	Success depends on quality data and human-in-loop.	• Fast multi-objective optimization • Balances potency, ADMET	• Retrospective validation •Reproducibility issues
3	**Generative AI in drug discovery (2024)**	Amit Gangwal & Antonio Lavecchia	Deep generative models for molecular design.	Promising but requires feedback loops and reward tuning.	• Explores chemical space • Rapid molecule design	• Synthetic infeasibility • Dependence on validation
4	**AI in discovery & clinical design (2025)**	Alberto Ocana et al.	AI aids target ID, structure prediction, trial design.	Needs high-quality data and ethical frameworks.	• Better target discovery • Smarter trial design	• Dataset bias • Regulatory hurdles
5	**AI reshaping drug development (2025)**	Sarfaraz K. Niazi & Zamara Mariam	AI in design, ADMET prediction, clinical development.	Cuts cost/time but faces bias and transparency issues.	• Faster candidate selection • Predicts ADMET	• Opaque models • No fully approved AI drugs

III. EXISTING VS PROPOSED SYSTEM:

Aspect	Existing System	Proposed System
Drug Discovery Approach	Uses ML, DL, GANs, and Transformers separately	Integrates Transformers with patient profiling into a single framework
Focus Area	Primarily molecular-level predictions (drug–target, protein folding, molecule design)	Combines molecular-level predictions with patient-level data (EHRs, genetics, health history)
Data Source	Chemical libraries, biomedical literature, lab experiments	Combines chemical data with Electronic Health Records (EHRs)
Accuracy	High for molecular predictions but less reliable in clinical outcomes	High molecular accuracy plus better clinical relevance
Limitations	Ignores patient variability; higher risk of clinical trial failure	Personalized approach reduces trial failures and speeds up drug approval
Healthcare Application	General drug design, drug repurposing, protein modeling	Personalized medicine and clinical decision support

IV. METHODOLOGY:

The proposed framework integrates transformer-based molecular modeling with patient profiling drawn from electronic health records (EHR) to advance drug discovery and personalize provision of health care. The enlargementc of the framework is self-possessed of the steps outlined below.

Step 1: Data Gathering

The implementation of the framework begins with the process of compilation of two modalities of data. From the drug side, drug molecules, protein sequences, and chemical compounds are compiled from reputable resources such as PubChem, ChEMBL, and the Protein Data Bank (PDB). From the patient side, de-identified EHRs can contain useful insight as it relates to medical history, genetics, demographics, and history of previous treatments.

Step 2: Data Preprocessing

Once the data is gathered from the previous step, the next step encompasses data organization. Molecules are reformulated as representations to be made interpretable by AI models including but not limited to SMILES strings, molecular graphs, or protein sequences. Similarly, the patient data from the EHRs must be cleansed and structured to provide useful perception regarding comorbidities, genetic markers, and treatment results.

Step 3: Modelling ofMolecular with Transformers

Now to incorporate advanced artificial intelligence models.

BioBERT conducts a search for valuable insight in biomedical texts with applications for drug-disease associations and drug and protein associations.
ChemBERTa can infer molecular properties through chemical structure analysis.
AlphaFold can model protein 3D structure, allowing investigators to identify binding sites and predict the mechanism of action of a drug.

Step 4: Patient Profiling and Integration

With the help of machine learning, patient information will be processed and risk factors, gene variations, and past drug usage patterns will be identified. This information will be subsequently tied to the predictions of the molecules to identify the best candidate drug(s) likely to act in a favorable manner for particular patient groups.

Step 5: Integrate Framework

In this step, outputs from the transformer models and those from the patient profiles are integrated. There is a reinforcement learning layer to optimize candidate drugs to best balance efficacy, safety, and synthetic ease.

Step 6: Outcome Generation

Lastly, the framework will output the best candidate drug(s) optimized for the patient-specific criteria. The results will then be communicated to a clinician through a decision-support system and be more effective and personalized treatment recommendations.

V. IMPLEMENTATION DETAILS:

The proposed framework is implemented using a combination of modern technologies, data sources, and programming environments capable of performing both molecular modeling and patient profiling.

Tools and Frameworks

Python is the foundational programming language, as it has support for both AI and machine learning. TensorFlow and PyTorch are employed for model development, as they enable deep learning, while the Hugging Face Transformers library is employed to run BioBERT and ChemBERTa. DeepMind's AlphaFold is used to predict protein structures; in coordination with Scikit-learn for patient profiling/processing. Pandas and NumPy will be used to handle and analyze the data

Datasets

Two wide categories of datasets will be required:

Molecular data:

- PubChem for chemical compounds and molecular properties.
- ChEMBL for bioactivity data and drug-target interactions.
- The Protein Data Bank (PDB) which provides 3D protein structures

Patient data:

- Electronic Health Records (EHRs) which give masked patient data such as medical history, genetic markers, demographics, and treatment results.

Implementation Workflow

The pipeline begins by collecting both patient and molecular data from the mentioned sources; after this, the data is preprocessed and structured. For instance, the molecules can be represented as SMILES strings, molecular graphs, or protein sequences, while EHR data is cleaned and formatted to highlight the most important features, for example, genetic markers, and comorbidities. Then, transformer models are used to process the data, data and produce well-defined profiles. The two outputs are then combined, along with a reinforcement learning layer for drug candidate optimization while balancing efficacy, safety, and synthetic viability. Finally, a including applications like BioBERT to extract biomedical knowledge, ChemBERTa for predicting chemical properties, or AlphaFold for predicting 3D protein structure. Simultaneously, machine learning models are used to analyze patient clinical decision support interface provides the developed recommendations, individualized for healthcare practitioners and patients.

Case Study: Cancer Research

To illustrate this approach an example of a cancer research case study could be used. For instance BioBERT makes drug–protein predictions based on the outcomes of a search of the cancer literature, ChemBERTa provides anti-cancer activity in compounds, and AlphaFold makes predictions about protein structures relevant to cancer biology (eg, tumor suppressors).

Patient profiling from EHRs also identifies genetic markers and combinations of responses to treatment as patients with cancer experience multiple modalities over time. This approach integrates together both the functional information on drugs that emerge from the artificial intelligence programs with profiling to recommend drug

candidates at a molecular level and ultimately applying to targeted patient populations to improve clinical trial success rates and to develop personalized oncology therapies.

VI. RESULTS AND DISCUSSION:

The proposed framework is expected to make drug discovery faster, more accurate, and more clinically relevant. By combining transformer-based molecular modeling with patient profiling from EHRs, it can identify promising drug candidates with greater precision and reduce the chances of trial failures. Unlike existant that focus mainly on molecules, this outlook includes patient-specific factors, ensuring that the selected drugs are productive for real-world populations.

Along with accuracy, the framework improves efficiency by minimizing dependence on trial-and-error methods. AI-driven predictions empowering quicker screening, and reinforcement learning ensures only the most productive, safe, and feasible candidates are emphasize. Compared to existing approaches, this dual-level integration makes the framework more robust and positions it as a powerful solution for personalized medicine and clinical decision support.

Current AI methods in drug discovery show significant accuracy at the molecular level, however, they do not perform as well in the clinic where variability is presented by patients. The proposed framework provides a solution to this issue by obtaining high accuracy at both the molecular and patient level by integrating molecular predictions with a patient profile using electronic health records (EHRs).

In terms of speed, current methods are faster than conventional discovery but would still go through multiple steps of validation. The proposed framework aims to accelerate discovery phases further by combining transformer based models via patient profile which can assist with quicker and more reliable candidate identification.

Many current methods have low clinical trial success rates due to ignoring genetic, demographic, and patient-specific variables. The proposed framework combines this patient data earlier in the pipeline with the expectation to improve overall trial success rates.

In terms of resource utilization, an issue with traditional AI systems is wasting resources due to low drug–patient compatibility in a preclinical and clinical model. The new framework will optimize resources by identifying and shortlisting only safe, effective, and synthesis prepared candidates using a reinforcement learning model prior to moving into a later stage.

Lastly, in the context of healthcare impact, current systems mostly emphasize general drug design and repurposing, rather than personalization. This is in contrast to the presented framework, which offers personalized medicine and an AI-enabled, real-time decision support tool to aid physicians in tailoring treatment to each patient.

VII. CHALLENGES AND LIMITATIONS:

Regardless of the enormous potential of the proposed framework, a couple of challenges still remain to be fulfilled for effective real-world implementation.

• Data Privacy in EHRs

Patient health records are immensely sensitive material and guarding it an important problem. Routinely, strict privacy, security, and legal compliance to regulations such as HIPAA and GDPR are needed before EHR data can be safely employed for AI research.

• High Computational Cost

Complex AI models, especially transformers and models based on reinforcement learning that utilize only the latest in terms of high-performance GPUs and storage. This can become expensive and complex leaving few resources for smaller research teams or healthcare systems.

• Ethical and Regulatory Issues

The use of AI for healthcare selection carries proper concerns such as bias, transparency and accountability issues. Moreover, seeking regulatory approval for AI-based drugs or clinical intervention systems can be cumbersome and slow

down the actual applications in hospitals and research laboratories..

VIII. FUTURE WORK:

Shortly, the framework could be expanded and enhanced in a variety of areas:

Including Multi-Model Data

Alongside EHR data, incorporation of genomics, imaging, and health wearables data could provide an even more holistic patient health profile resulting in improved predictive performance and more personalized treatments.

Increasing Model Transparency

Increased transparency and interpretability with AI models will make it easier for providers to access how predictions are formulated. This could not only build trust but also allow for safer and more reliable use of AI.

Clinical Testing in the Real World

To fully validate the framework, large-scale clinical trials and real-world testing is necessary to validate its performance and accuracy before it is broadly adopted in hospitals and research space.

IX. CONCLUSION:

The present research showcases the transformative potential of Artificial Intelligence applications in drug discovery and Health Care. By bridging molecular prediction via transformer molecular modeling to patient profiled by electronic health records, the developed platform has the promise of better accuracy, faster identification of drug candidates, and lower risk for clinical trial failure. This platform is unique and does not end with molecular prediction as is common with traditional approaches, but it adds a patient specific component, which gives results more clinical relevance. In the end, this brings health care closer to personalized medicine, where treatment is tailored to the specific characteristics of an individual.

X. ACKNOWLEDGEMENTS

The authors acknowledge the utmost appreciation to their guide and mentor for the ample guidance, motivation, and constructive suggestions provided during the completion of this research work. The authors express their gratitude to the Department of Computer Science faculty members for their continuous support and providing the necessary facilities and environment to help facilitate this research. The authors would especially like to thank their peers and researchers for their contributions in the field of Artificial Intelligence and Drug Discovery, and prior work conducted in this area that helped anchor and strengthen the authors' contributions. Finally, the authors would like to acknowledge their family and friends for their ongoing motivation and understanding during the research.

XI. REFERENCES:

1. J. Smith et al. (2023). Predicting molecular properties in drug discovery using deep learning.
2. Johnson et al. (2024). Generative adversarial network for new compound development.
3. A. Patel et al. (2025). Use of AI Powered Clinical Decision Support systems in healthcare.
4. M. W. Mullowney et al. (2023). AI and machine learning are used for joining with cheminformatics and omics to discover naturalist product drugs.
5. Oprea, T. I., & Hasselgren, C. (2024). Artificial Intelligence in Drug learning Process - Opportunities and Challenges.
6.Lavecchia, A. & Gangwal, A. (2024). Drug Discovery Using Generative AI: From Models To Molecules.
7. A. Ocana et al. (2025). Discovery cancer research by integrating AI into early clinical design and discovery, A revolution.
8. Mariam, Z. & S. K. Niazi (2025). AI is transforming the development of drugs from design to clinical application.
9. Li, X., Zhang, S., and Chen, J. (2024). Analysis of the use of AI in drug development.
10. Wang, Y., and others (2023). Graph neural networks for molecular representation learning in drug discovery.

11. Zhou, T., and Kim, D. (2024). Dispersion models for structure-based drug plan and the blend of new molecules.

