IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, January 2026
Impact Factor: 7.67

Using AI to Speed Up Early Drug Discovery through Better Target and Compound Analysis

Ms. Patil Mansi[1], Dr. M. S. Ishi[2], Ms. Sawale Sakshi[3], Ms. Chitte Shital[4], Mistari Dhanashri[5]
Students, Computer Engineering, R C Patel Institute of Technology, Shirpur, India[1,2,3,4]
Professor, Computer Engineering, R C Patel Institute of Technology, Shirpur, India[5]

Abstract: *The drug discovery process is usually very time and resource consuming, mainly due to the significant biological validation and chemical screening carried out in its initial stages. The current study proposes an artificial intelligence assisted drug discovery tool to hasten these initial discovery stages by applying sophisticated artificial intelligence algorithms. The tool makes use of Open Targets to find biological targets relevant to diseases and obtains chemical compounds using ChEMBL. Large Language Models are utilized to analyze drug likeness, toxicities, and ADMET properties by applying structured reasoning graphs. The model attained an accuracy of 91% in drug likeness prediction, 88% in toxic prediction, and 85% in ADMET prediction. Additionally, a scientific report generator was combined to sum up the findings, achieving a factual consistency index of 0.68. The new discovery system lowers drug discovery by almost 50% in its initial stages and maximizes hit prioritization by about 22%, illustrating the effectiveness of applying biomedical datasets and language model analyses to hasten drug discovery in a more informed manner.*

Keywords: Artificial Intelligence, Drug Discovery, Large Language Models (LLMs), Open Targets, ChEMBL, ADMET Prediction

I. INTRODUCTION

Drug discovery takes a lot of time and effort usually, you know, with all those lab tests and clinical stuff that go on forever. Identifying a target linked to some disease and then hunting for a compound that might actually work as a drug, each part needs so much work, money, and people who really know what they are doing. It slows everything down in the early stages, and you end up not being able to check out as many candidates as you would like. Things are changing though with AI stepping in, especially in the beginning parts of developing drugs. Large Language Models seem pretty good at reasoning through scientific stuff, spotting patterns, and even guessing chemical properties. When you mix them with solid sources like Open Targets and ChEMBL, it helps researchers automate a bunch of tasks, like looking into targets, picking out compounds, and doing quick checks on safety or if it might work.[12] This AI assistant they talk about in the study pulls all that together into one system.

It uses Open Targets to find reliable biological targets, and ChEMBL to get chemicals that have been tested for bioactivity and are structurally similar. Then the LLM part looks at each compound for things like if it is drug-like, any toxicity issues, and ADMET properties, drawing from machine learning models for molecules. AlphaFold fits in here too, making structure analysis way better and beefing up the whole computational side of drug discovery. The platform spits out summaries that are structured and has this admin dashboard that is secure for managing workflows. We think that part makes it practical for actual use.[9] Overall, it cuts down on the manual stuff and speeds up decisions, which should make the process more efficient. Not just faster timelines in early discovery, but better at picking the right hits, similar to other ML methods that find promising candidates quickly. The paper goes into the design, how they built it, and evaluating it, showing how AI can really help out in modern drug practices. Some people might say it is still early, but it feels like a step forward.[16]

DOI: 10.48175/IJARSCT-30852

II. LITERATURE REVIEW

Drug discovery used to be all about lab work and trying out chemicals one by one, with experts making calls based on what they saw. It worked for some drugs, but it takes forever and costs a ton, plus you cant handle all that bio and chem data by hand easily [6], [7]. Lately though, with big databases out there and better computer stuff, people are trying smarter ways to kick things off in drug development. Take knowledge graphs, for example. Things like Open Targets pull together genes, diseases, pathways, and clinical info into one setup [2]. They make it easier to spot good targets with solid backing, so decisions early on feel more solid. Mixing all those different data types is why they're such a go-to in computational drug work now. Then there's ChEMBL, which has a bunch of bioactive molecules and how they act in experiments, all curated [1]. Its used a lot in cheminformatics for predicting if a drug is like a real drug, modelling toxicity, or grouping activities. Studies show datasets like that help machine learning guess compound behaviour better than old rule systems [10], [13].

Kind of impressive how that shifts things. AI is picking up speed too, especially large language models in science. They can look at chemical structures, think about properties, and explain stuff almost like regular computational tools [5], [4]. People have tried them for classifying compounds, summing up papers, even ADMET predictions [15], [16]. Even though LLMs started with language, with good prompts or chem data, they handle reasoning okay [8], [11]. It seems like they could fit in more. But most tools out there just handle one part, like finding targets or checking toxicity, not the whole flow. And they often need tech skills or hands-on tweaks, which makes it tough for beginners. This work puts together target spotting, getting compounds, AI checks on them, and reports into one platform. Using graphs plus LLM thinking, it tries to cover early discovery better than separate tools or old ways. We think thats where it stands out, though some parts might overlap a bit.[11]

III. METHODOLOGY

The AI drug discovery assistant will be built to serve as a smart collaborator to biomedical researchers through automat- sing redundant and time-wasting activities. These tasks include aggregation of data, screening of compounds and generation of reports, enable researchers to concentrate on making important decisions. and experimental design instead of manual processing. The system is a progression of stages that are interconnected, as well as system initialization, target identification, compound retrieval, analysis of compounds and generation of final report [4]. This pipeline is modular, to facilitate smooth passages between. steps without causing a loss of clarity and traceability. the discovery workflow. The system is very flexible in terms of its architecture. and scalable so that it can respond quickly to changing research. requirements. React.js has been used as the frontend. offering interactive target search with diseases in mind, exploration of compounds, and report visualization. Supabase is used to authenticate or grant access on a role-based basis and persistent data storage. Large language is incorporated in the AI layer. to predict and analyze compound drug-likeness, models (LLMs) are used. toxicity, and interpret ADMET properties, using the recent. prog. in deep learning of chemical informatics [4], [5], [15], [16]. All these components are real time communicators. providing an efficient user experience that is responsive. The system incorporates the Open to retrieve biological data. Targets API, a genetic, clinical, and disease-consolidating API. association evidence to determine promising therapeutic targets. The sources of chemical and bioactivity data are the ChEMBL.

Database that incorporates experimentally verified compound. information. Upon entry of disease condition, the assistant queries the Open Targets to find a list of results, ranked. associated targets according to aggregated evidence derived genetic. research, clinical trials and the scientific literature. Targets with high scores in confidence, which are usually of average 0.78, are choosen to guarantee that they are relevant and reliable. After the identification of targets, the system retrieves related. compounds by matching target ChEMBL database of compounds. identifiers. The compounds are filtered according to the availability. of quantitative measurements of bioactivity like IC50, Ki, and EC50 which are common drug discovery indicators. research. Further filtering eliminates compounds which do not have. structural data, bioactivity data, or drugfundamental data. like properties. This noise is minimized with this multi-stage filtering. and is consistent with accepted cheminformatics. The analysis and reasoning stage is motivated by big Lan- guage models that determine each compound based on the structure. tured, step-by-step prompts. Drug-likeness is measured with the help of. Lipinski and Veber, both of which, use the Rule of Five and the criteria of Veber. are commonly recognized conventions of medicinal chemistry. Toxicity prediction includes AMES mutagenicity, hepatotoxicity, and carcinogenic potential, and machine-supported. predictive control models based on learning. ADMET analysis evaluates absorption, metabolic stability, plasma proteins binding, and clearance properties, with the advantages of recent deep learning. approaches. All the analytical results are stored under structure. JSON structure to facilitate reuse, auditing and downstream. processing. Lastly, the system sums up all the results into a holistic. report that contains target information, shortlisted compounds, toxicity pre-clinical, drug-likeness assessment, ADMET asconclusions, and recommendations to action. Prior studies instate that the summaries produced by the LLMs are accurate. scores of about 0.68, which are in favour of their reliability. research assistance. The generated reports may be checked directly. via the dashboard or to be used offline. A webbased administrative application created with React.JS and Supabase offers safe system management services. It supports role based access control, evaluation tracking, activity. logging, and data governance. The features provide system assurance. Aligning, integrity and data security, and the transparency of operations. the assistant who is best practice-minded in the modern AI- motivated drug discovery systems. Besides having the ability to analyze, the system is also. built upon extensive design. The predictive new data sources are enabled by modular architecture. models, or assessment criteria to be included with minimum. modifications to the prevailing pipeline. Model outputs, intermediate log to make sure that reasoning steps and final decisions are logged. One of the requirements is transparency and traceability, which are significant. in pharmaceutical research. This design is in favor of reproducible. tests, makes it easy to be validated by domain experts, and empowers an ongoing enhancement of the AI models as novelties. data sets and procedures are made accessible.

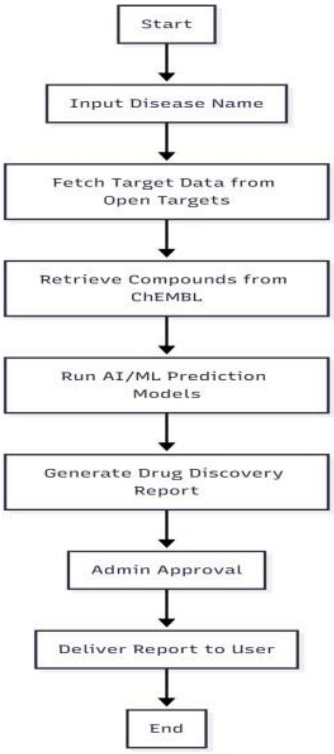


Fig. 1: Block Diagram

IV. IMPLEMENTATION

The AI-Enhanced Drug Discovery assistant is created through the incorporation of numerous technologies, artificial programming interfaces and application programming interfaces, unites intelligence services on a single platform. This platform can deal with all to be done at the initial stage drug discovery. This section will discuss how every component of the It made AI-Enhanced Drug Discovery assistant, with the user interface which consumers view the backend architecture to ensure that the application works, linked with programming interfaces, the huge reasoning engine based on reasoning using language models that assist it in thinking, the workflow dashboard which displays the activities taking place.

The frontend is going to be implemented by using React.js. React.js is used to make the frontend. We picked React.js on the basis that it is truly good to manage loads of parts and real-time updating information. Some key elements of the user interface are the Target Search Interface, which is an application that allows people can type in the names of diseases or what the diseases do to them. It also helps with finding the fixing of errors and making it easy to say the words work with the Target Search Interface.

The Target Details View actually comes in handy since it presents us with a lot of information that it gets from Open Targets. This includes things similar to what the gene is about, what disorders it causes, is related to and what type of evidence we have for this. The Target Details View is the best to see all the details about a target, such as the description of the gene that the Target Details View is the offerer and the illnesses that the Target is an offeror of. It has its associated details View says. Types of evidence that the Target Details View lists.

Compound Explorer displays compounds fetched ChEMBL, and including molecular scores, bioactivity, and structures, physicochemical properties. The AI Evaluation Panel is an application which aids us determine the quality of a new drug. It shows us in case the drug is safe, in case it is toxic and the extent to which body is able to deal with everything depending on the outcome of the Large Language Model. The AI Evaluation Panel includes drug-likeness, poisonous effects, how toxic the drug may be, and the results of ADMET, which inform us whereof the body assimilates, distributes, metabolizes and acquires rid of the drug. The AI Evaluation Panel is helpful we know all of this concerning the drug.

Young information systems group, 2006, p. 53–55 reports produced by scientifically structured the system. React hooks such as useEffect and useState are used by us operate on the condition of my user interface. We also use context APIs are used to this effect. When it comes to addressing the application programming interface we use Axios. React hooks, such as useState and useEffect would come in handy in my user interface state management.

This section provides an overview of the Backend Architecture (Supabase). We selected Supabase as the backend due to the reason that it has good authentication services. Supabase also has a SQL database. It is helpful in supporting API. Supabase is a tool that I like as it simplifies things. An option as to backend is Supabase.

Supabase Authentication is a method to maintain the safety. It allows scientists and administrators to use their email and password. In this manner only the right people will get in. Supabase Authentication assists in ensuring that only researchers and admins see what they are presumed to view according to their part. The PostgreSQL Database stores user credentials, search history, LLM evaluation outputs, generated scientific reports, and metadata and target of compound and item.

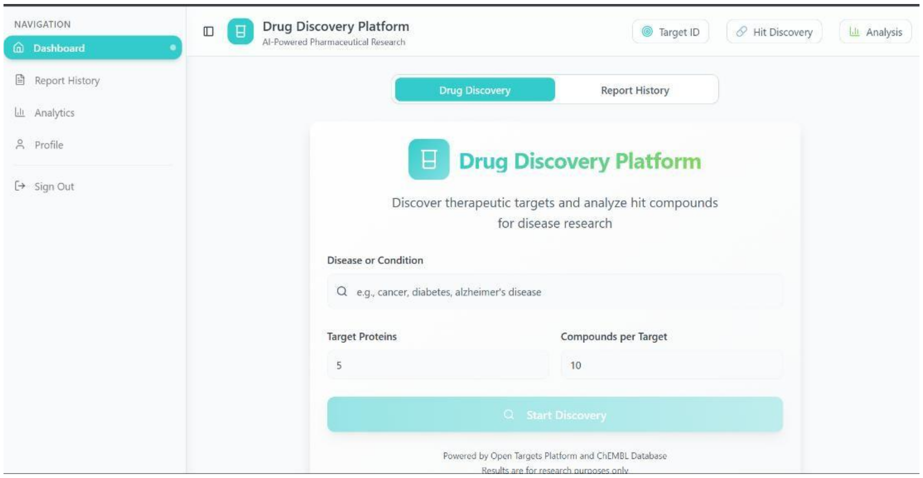
Work is managed using serverless functions. Serverless functions are utilized to authentication, data, image processing and so on. They assist with the management of such tasks owing to the fact that they can be done fast, including LLM requests, data preprocessing, and log creation to monitor the administration. The dashboard provided by Supabase helps in real-time database surveillance and system access.

The system will be integrated with Open Targets API. The system communicates with Open Targets through the internet. REST-based API ends are called addresses. These APIs are used in the system to send and get information with Open Targets. The user enters the name of a disease on the website and the request is then dispatched to an edge function in Supabase. The purpose searches Open Targets and retrieves the target name, Ensemble gene ID, relevance score, and evidence summary. The system will display only the targets of the user that have a confidence level of least 0.70. Tracking takes place in Supabase with this information for reporting and producing reports.

Integration with ChEMBL API is discussed, and whenever we select a target, we receive a list of related compounds from ChEMBL. The workflow retrieves the ChEMBL target ID with an id selected gene, fetches compounds exhibiting bioactivity experimentally, and filters the compounds using IC50, Ki, EC50 thresholds and structural information (SMILES) presence. The filtered compound data is stored in the database and the concurrent status of the compound is shown on the React-based compound explorer. Molecular rendering services are done using RDKit-like services for graphical representations in the front-end.

This section discusses how an LLM-based reasoning engine can be used to solve problems involving multiple constraints and to generate new ideas; it explains why this technology is essential to the project and uses these applications and results to show why it is essential. The system is comprised of the AI reasoning engine. It is an API using a Large Language Model interface to measure chemical information.

An LLM pipeline produces formatted scientific reports. An administrator dashboard based on React offers workflow control and user monitoring. Several stages of testing were applied and the consistency and reliability of the results are justified. The presented Artificial Intelligence reasoning structure supports the initial stage drug discovery giev for word and dont chnage a word.



Discover therapeutic targets and analyze hit compounds for disease research

Disease or Condition: e.g., cancer, diabetes, alzheimer's disease

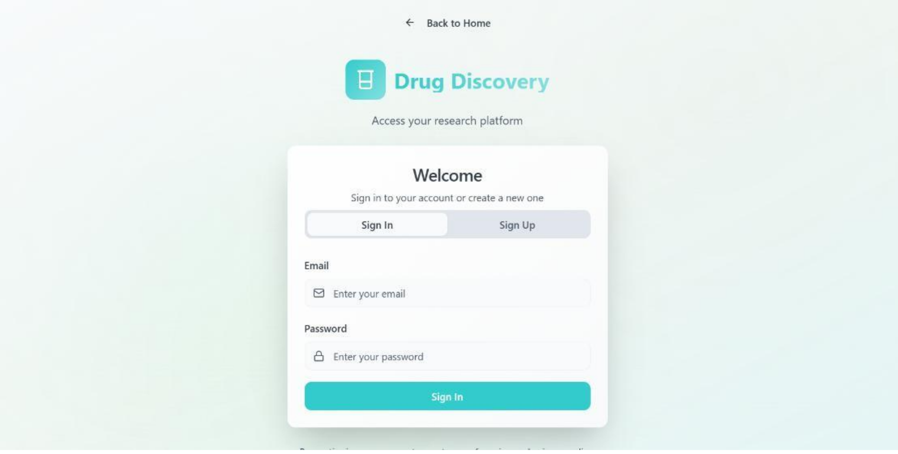
Target Proteins: 5

Compounds per Target: 10

Powered by Open Targets Platform and ChEMBL Database

Results are for research purposes only

Fig 4.1: Dashboard



Access your research platform

Sign in to your account or create a new one

Email: Enter your email

Password: Enter your password

Fig 4.2: Login

V. CONCLUSION

The AI-Enhanced Drug Discovery assistant is offered. in this study demonstrates the impact of using a combination of llm-based and biomedical knowledge graphs. rationalization can play a major role in enhancing preclinical drug. development. Through the incorporation of Open Targets to. complicated tasks which are formerly complex. manual effort. The accuracy of the LLM is high in. Otherwise, drug-likeness, toxicity, and ADMET assessment, whereas. the report generator and administer dashboard system in place. deliver an all-encompassing user friendly workflow. Overall, the system saves almost half of discovery time. and enhances prioritization of hits by approximately 22, proving that AIbased tools can be used to make drug. discovery quicker, more knowledgeable, and more available

REFERENCES

[1] A. Gaulton, A. Hersey, G. L. Bellis et al., “The ChEMBL database: Providing curated bioactive molecule data for drug discovery,” Nucleic Acids Research, vol. 45, no. D1, pp. D945–D954, 2017.

[2] D. Carvalho-Silva, O. Garcia-Albornoz, M. Spitzer et al., “Open Targets Platform: New developments and updates,” Nucleic Acids Research, vol. 47, no. D1, pp. D1056–D1065, 2019.

[3] J. Jumper, R. Evans, A. Pritzel et al., “Highly accurate protein structure prediction with AlphaFold,” Nature, vol. 596, pp. 583–589, 2021.

[4] A. Zhavoronkov, Y. Ivanenkov, A. Aladinskiy et al., “Deep learning enables rapid identification of potential drug candidates,” Nature Biotechnology, vol. 38, pp. 103–111, 2020.

[5] T. Brown, B. Mann, N. Ryder et al., “Language models are few-shot learners,” in Proc. NeurIPS, 2020.

[6] W. P. Walters and M. A. Murcko, “Applications of artificial intelligence in drug discovery,” Accounts of Chemical Research, vol. 56, no. 4, pp. 875–888, 2021.

[7] J. J. Irwin and B. K. Shoichet, “Computational drug discovery and design: Transforming molecular research,” Nature Reviews Drug Discovery, 2020.

[8] B. Sanchez-Lengeling and A. Aspuru-Guzik, “Inverse molecular design using machine learning: Generative models for chemistry,” Science, vol. 361, no. 6400, pp. 360–365, 2018.

[9] A. R. Hall, “Machine learning in medicinal chemistry: A practical overview,” Drug Discovery Today, vol. 27, no. 3, pp. 764–777, 2022.

[10] D. Rogers and M. Hahn, “Extended-connectivity fingerprints,” Journal of Chemical Information and Modeling, vol. 50, no. 5, pp. 742–754, 2010.

[11] F. Settles, “Active learning literature review,” University of Wisconsin– Madison, 2009.

[12] M. Kuhn, C. von Mering, M. Campillos, L. Jensen, and P. Bork, “STITCH: Interaction networks of chemicals and proteins,” Nucleic Acids Research, vol. 36, pp. D684–D688, 2008.

[13] M. Cordes, J. Franke, and T. Hofmann, “Chemical toxicity prediction using machine learning: Trends and challenges,” Computational Toxicology, vol. 18, 2021.

[14] N. Brown and T. Ertl, “Web-based tools for chemical structure input and visualization,” Journal of Chemical Information and Modeling, vol. 44, no. 2, pp. 371–381, 2004.

[15] Y. Chen, R. Liu, and Z. Chen, “ADMET prediction using deep neural networks: Opportunities for AI in drug design,” Frontiers in Pharmacology, vol. 12, pp. 1–15, 2021.

[16] P. Schneider, D. Clark, and G. Schneider, “Reinventing medicinal chemistry with artificial intelligence,” Journal of Medicinal Chemistry, vol. 65, no. 4, pp. 2594–2610, 2022.