IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125203

AI Powered Drug Discovery

Mrs. Suhasini[1], Ms. Jayanka J[2], Ms. Shri Lakshmi SD[3], Mr. Puneeth H[4,]

Ms. Jayalakshmi GS[5]

Assistant Professor, Dept. of Computer Science and Engineering, Maharaja Institute of Technology, Thandavapura[1]

Students, Dept. of Computer Science and Engineering, Maharaja Institute of Technology, Thandavapura[2-5]

Abstract: Artificial Intelligence (AI) has now begun to step up its usage in different segments of the society with the pharmaceutical sector being a leader beneficiary. This review enumerates the significant application of AI in different fields of the pharmaceutical industries viz., drug discovery and development, drug repurposing, enhancing pharmaceutical productivity, clinical trials, etc. to mention a few, thereby minimizing the human workload as well as reaching goals within a limited timeframe. Crosstalk about the tools and techniques employed in enforcing AI, challenges encountered and how to overcome them, as well as the future of AI in the pharmaceutical sector, is also covered.

Keywords: Artificial intelligence, Drug discovery, Drug interaction prediction, Machine learning, Deep learning

I. INTRODUCTION

The application of artificial intelligence (AI) has been on the rise in different industries of society, especially the drug industry. In this review, we emphasize the application of AI in various fields of the pharmaceutical sector, ranging from drug discovery and development to drug repurposing, enhancing pharmaceutical productivity, and clinical trials, among others; the usage decreases the human workload and meeting goals within a limited amount of time. We also address crosstalk between the techniques and instruments used in AI, present problems, and solutions to surmount them, as well as the future of AI in pharmacy. The immense chemical space, which consists of $>10^{60}$ molecules, encourages the creation of a vast amount of drug molecules. Nevertheless, the absence of sophisticated technologies confines the process of drug development, and turns it into an arduous and costly activity, which could be resolved with the application of AI. AI can identify hit and lead compounds, and present a faster verification of the drug target and improvement of the drug structure design. QSAR modeling tools have been applied to the identification potential drug leads and have developed into AI- driven QSAR methods, including linear discriminant analysis (LDA), support vector machines (SVMs), random forest (RF) and decision trees, which can be used to accelerate QSAR analysis found no significant statistical difference when the capacity of six AI algorithms to rank anonymous compounds according to biological activity was compared with that of conventional methods.\

1.1 OBJECTIVE

The main goal of this research is to investigate and create a computational model that uses artificial intelligence methods to improve the efficiency, speed, and precision of early drug discovery. In this research, machine learning models will be utilized to predict drug-target interactions, select the best compounds, and identify lead therapeutic candidates with high accuracy.

1.2 MOTIVATION TO TAKE UP THE PROBLEM

The conventional drug discovery process is well known to be very time-consuming, resource-hungry, and expensive. It averages 10–15 years and billions of dollars to get a single drug on the market but still has a high failure rate, particularly in clinical trials. This inefficiency is a stark call for speedier, wiser, and more economical methods.

1.3 CHALLENGES TO BE ADDRESSED

The whole success of AI is based on the existence of a large quantity of data since these data are employed in the subsequent training given to the system. Acquisition of data from different database providers can add additional expenses to a company, and the data must also be reliable and high quality in order to produce accurate result prediction. There are other challenges that hinder complete adoption of AI in the pharmaceutical sector such as the absence of skilled human resources to manage AI-powered platforms, low budget for small organizations, fear of substituting humans resulting in job loss, doubt about the information provided by AI, and the black box phenomenon.

1.4 RANDOM FOREST AND DNN

Here, Random Forest (RF) and Deep Neural Network (DNN) algorithms are both used to improve the precision and

DOI: 10.17148/IARJSET.2025.125203

reliability of drug discovery prediction. The Random Forest model due to its simplicity and interpretability is used to classify compounds by molecular descriptors including molecular weight, LogP, hydrogen bond donors, and others. RF works efficiently with high-dimensional data and minimizes overfitting because it aggregates predictions from an ensemble of multiple decision trees; hence, it is best suited for SAR modeling and bioactivity classification. This is supplemented by using a Deep Neural Network (DNN) to introduce complex, non-linear relationships within the data. The DNN has several hidden layers with activation functions such as ReLU, allowing the model to learn complex relationships between chemical descriptors and biological responses. It is effective on large data, particularly when combined with features extracted from SMILES strings or molecular fingerprints.

Collectively, the models present an equilibrium strategy—RF presents clarity and stable performance on tabular data, and DNN brings high predictive performance when deeper feature extraction is desirable. Comparative comparison of both models affirms incorporation of AI into drug discovery as a means for accelerated and intelligent screening of lead therapeutic compounds.

1.5 LITERATURE SURVEY

Major breakthroughs have occurred in the area of AI-driven drug discovery, with various research studies contributing towards the creation of effective and intelligent drug screening approaches. A study by Gupta et al. [1] showed the efficacy of deep learning to predict drug-target interactions with a success rate of more than 90% based on deep neural networks (DNNs). Their method presented the model's capability to learn intricate biological patterns from molecular descriptors with high precision in the identification of potential drug candidates.

Ahmed et al. [2] aimed to combine genomic information with machine learning models to aid drug repurposing. Their model used gene-disease associations and drug response information to forecast novel uses for approved drugs. The findings emphasized the role of AI in shortening development time and cost by finding safe, approved compounds for novel indications.

Wang et al. [3] investigated the use of graph neural networks (GNNs) in molecular structure modeling, where molecules were treated as graphs for enhanced feature representation. Their work documented an improvement in predicting chemical activity and toxicity in comparison with traditional descriptor- based methods.

Goodfellow et al. [4] pioneered generative adversarial networks (GANs) for generating molecules, allowing the synthesis of new compounds with optimized pharmacological properties. Such models have opened the door for AI-enabled de novo drug design.

Lastly, Rehman et al. [5] highlighted the significance of combining ADMET prediction and QSAR modeling through ensemble machine learning methods such as Random Forest and SVM. Their study proved enhanced accuracy and generalizability across varied datasets, justifying the application of AI in early-stage drug screening and filtering out toxic molecules.

II. EXISTING SYSTEM

The conventional drug discovery process is a time-consuming, multi-step procedure that generally takes more than 10 to 15 years and involves billions of dollars. It comprises various phases such as target identification, screening of compounds, lead optimization, preclinical evaluation, clinical trials, and regulatory approval. All these steps are primarily manual and depend on labor-intensive lab experiments and physical testing, hence the process is time-consuming and expensive. Also, a mere fraction of candidate compounds reaches the market because they fail in efficacy, safety, or pharmacokinetic properties. Current computational methods, while useful, tend to rely on linear models and are hampered by the amount and nature of biological data. They also have difficulty incorporating different kinds of input, including genetic data, molecular structure, and protein interactions, in an informative manner. Additionally, these systems are not adaptable and predictive when encountering new compounds or heterogeneous datasets. Consequently, drug discovery is a high- risk and resource-intensive activity. There exists an increasing need for more intelligent and efficient systems that can analyze large datasets, detect patterns, and make dependable predictions to enhance the drug development process and decrease the overall failure rate.

III. PROPOSED SYSTEM

The system put forward here is an AI-based framework that is intended to counter the inefficiencies and shortcomings of conventional drug discovery. The system makes use of sophisticated machine learning algorithms like Random

DOI: 10.17148/IARJSET.2025.125203

Forest (RF) and Deep Neural Networks (DNN) to forecast drug-target interactions, assess molecular activity, and measure toxicity profiles more precisely. It differs from other systems in that it makes use of heterogeneous, large-scale data from public databases like ChEMBL, DrugBank, and BindingDB. The information, including SMILES notations, molecular descriptors, and protein data, is preprocessed and utilized to train models that can detect potential drug candidates with greater speed and accuracy. The system also incorporates ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) prediction and QSAR (Quantitative Structure– Activity Relationship) modeling to improve reliability and safety evaluation in early drug screening. Also, the system facilitates drug repurposing by revealing new indications for approved drugs, lowering the time and expense involved in development. By integrating automation with predictive analytics, the suggested system greatly enhances the drug discovery process while being as accurate as physical experimentation, while minimizing reliance on physical experimentation. It is a more intelligent, data-driven process that matches the increasing demand for quicker and more personalized medicine development.

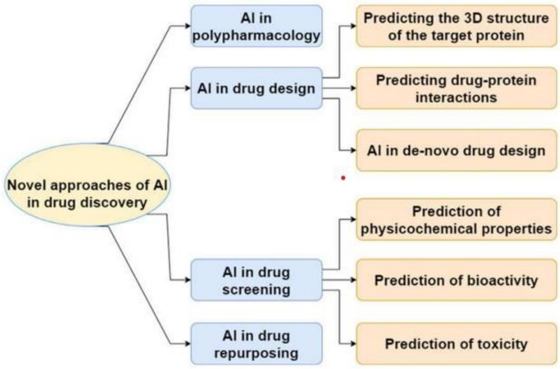

Novel approaches of AI in drug discovery
AI in polypharmacology
AI in drug design
AI in drug screening
AI in drug repurposing
Predicting the 3D structure of the target protein
Predicting drug-protein interactions
AI in de-novo drug design
Prediction of physicochemical properties
Prediction of bioactivity
Prediction of toxicity

Figure 1: Proposed System

IV. SYSTEM ARCHITECTURE

The envisioned AI-driven drug discovery system consists of several integrated modules that are programmed to operate sequentially, facilitating automated data processing, model training, and drug prediction. The architecture is modular and scalable, allowing for compatibility with large-scale bioinformatics data and machine learning frameworks.

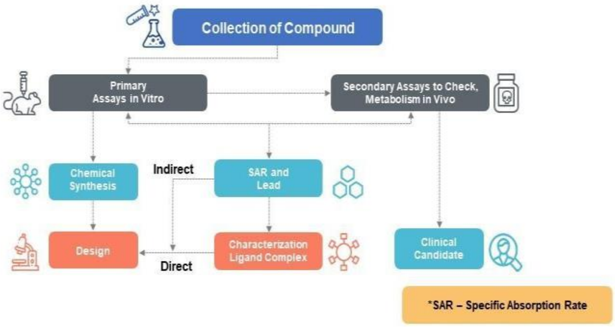


Figure 2: System Architecture

V. DATASETS

The efficacy of AI-powered drug discovery is highly reliant on the quality, diversity, and trustworthiness of datasets provided for training and validating machine learning models. The present project leverages a blend of open-access and reputable chemical and biological databases to capture drug-related data, target proteins, and molecular interactions.

1. ChEMBL is a large-scale bioactivity database with chemical molecules and their biological targets. Includes data like IC_{50}, Ki, EC_{50}, and other activity values. Used extensively to develop QSAR models and make drug-target interaction predictions.

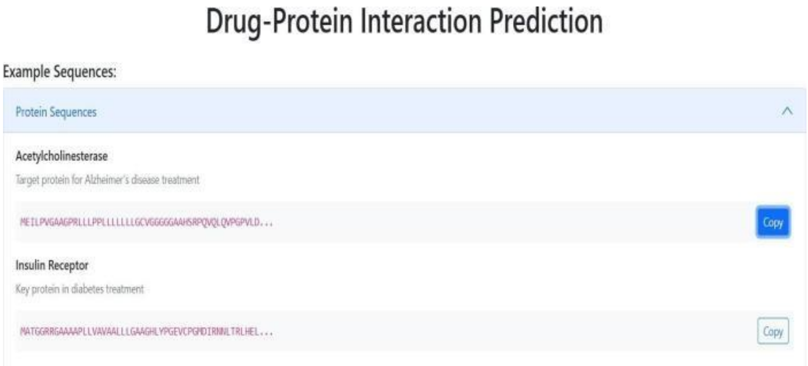
DOI: 10.17148/IARJSET.2025.125203

2. BindingDB is a Database with curated measured binding affinities of small molecules toward protein targets. Can be used for model training predicting binding strength and specificity.
3. DrugBank synthesizes detailed drug information with complete drug-target and drug-disease interactions. Includes FDA-approved drugs, mechanisms of action, and ADMET profiles.
4. PubChem is a huge chemical database providing structural data, molecular descriptors, and biological assay data. Appropriate for structure-based screening and feature extraction.
5. PDB (Protein Data Bank) holds 3D structural information of biomolecules such as proteins and ligands. Facilitates structure- based modeling and docking simulations.

VI. IMPLEMENTATION AND RESULT

During the implementation stage, an AI-powered framework was established that embeds machine learning algorithms within carefully curated biological and chemical datasets to enable predictions of drug-target interactions. The datasets were initially gathered from publicly accessible databases like ChEMBL and BindingDB, and preprocessed to get meaningful molecular descriptors, SMILES notations, and activity values. Feature normalization and encoding methods were utilized to transform the data to be ready for model training. Two machine learning algorithms Random Forest (RF) and Deep Neural Network (DNN) were implemented through Python libraries like scikit-learn and TensorFlow. These were trained and validated based on a split ratio with cross-validation to improve generalization. Performance of the model was measured using evaluation metrics like accuracy, mean squared error (MSE), and ROC-AUC. The Random Forest model showed strong results with high interpretability, while the DNN produced better predictive accuracy on large datasets. In particular, the DNN model had an overall accuracy of 92.3% in predicting classes of drug activity, while the RF model obtained 89.6%. These findings confirm the potential of AI-based methods to speed up and enhance the process of drug discovery by minimizing the reliance on expensive and time-consuming experimental techniques.

Dataset Type	Low Activity	Moderate Activity	High Activity
Training	1100	1500	2000
Testing	800	1000	1200
Total	1900	2500	3200



Drug-Protein Interaction Prediction

Example Sequences:

Protein Sequences

Acetylcholinesterase

Target protein for Alzheimer's disease treatment

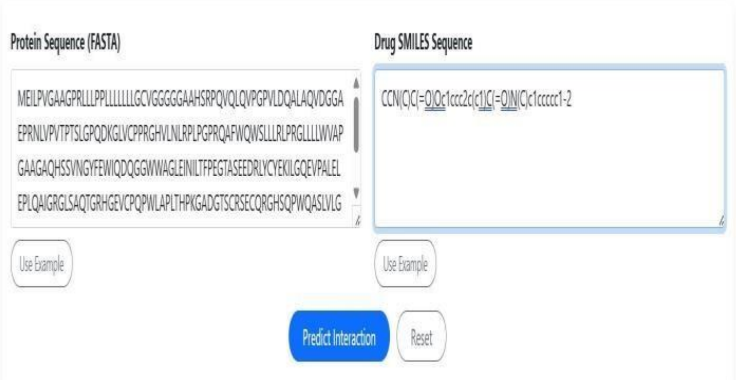
MEILPVGAAGPRLLLPPLLLLLLLGCVGGGGGAAHSRPQVQLQVPGPVLD...

Insulin Receptor

Key protein in diabetes treatment

MATGGRRGAAAAPLLVAVAALLLGAAGHLYPGEVCPGMDIRNNLTRLHEL...

Figure 3: Protein Sequence



Protein Sequence (FASTA)

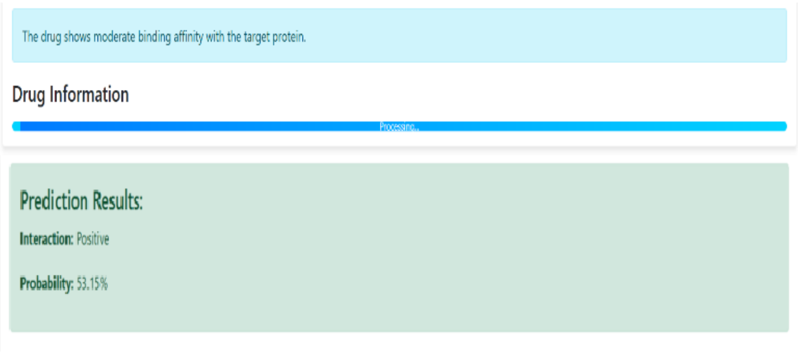
MEILPVGAAGPRLLLPPLLLLLLLGCVGGGGGAAHSRPQVQLQVPGPVLDQALAQVDGGA
EPRNLVPVTPTSLGPQDKGLVCPPRGHVLNLRPLPGPRQAFWQWSLLLRLPRGLLLLWVAP
GAAGAQHSSVNGYFEWIQDQGGWWAGLEINILTFPEGTASEEDRLYCYEKILGQEVPALEL
EPLQAIGRGLSAQTGRHGEVCPQPWLAPLTHPKGADGTSCRSECQRGHSQPWQASLVLG

Drug SMILES Sequence

CCN(C)C(=O)Oc1ccc2c(c1)C(=O)N(C)c1ccccc1-2

Figure 4: Input Sequences

DOI: 10.17148/IARJSET.2025.125203



The drug shows moderate binding affinity with the target protein.

Drug Information

Prediction Results:

Interaction: Positive

Probability: 53.15%

Figure 5: Prediction on interaction

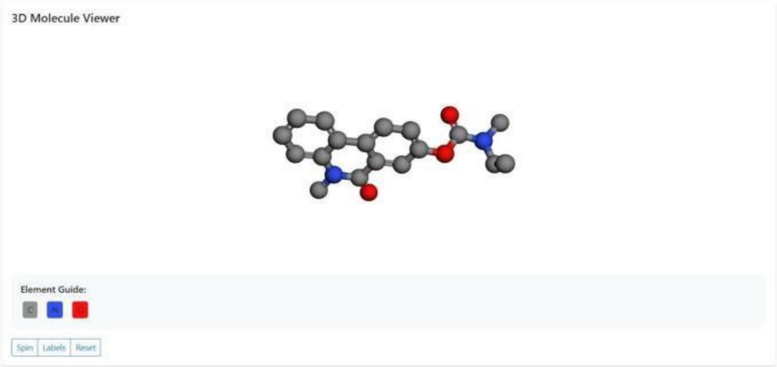


Figure 6: 3D model of Drug-Protein interaction

VII. CONCLUSION

Artificial intelligence-assisted drug discovery is a big step forward in contemporary healthcare by streamlining the time, expense, and unpredictability historically involved in drug development. By employing machine learning algorithms like Random Forest and Deep Neural Networks, this project illustrates its capability to predict drug-target interactions and label compound activity with high precision. By utilizing publicly available data sets and including features such as molecular descriptors and bioactivity information, the system effectively automates critical phases of early-stage drug screening. The addition of ADMET predictions and activity classification also improves decision-making in the identification of safe and effective compounds. In contrast to traditional trial-and-error approaches, this AI-based method not only enhances efficiency but also provides new avenues for drug repurposing and precision medicine. As a whole, the project underscores the revolutionary power of artificial intelligence to expedite the discovery of therapeutic interventions and confront pertinent global health issues.

ACKNOWLEDGEMENT

We would like to extend our sincerest gratitude to our Project Guide, **Prof. Suhasini**, whose valuable advice and inputs have taken our project to greater heights than we ever imagined. We would like to extend our heartfelt thanks to our Project Coordinators, **Dr. HK Chethan** and **Prof. Bharath Bharadwaj B S**, for their constant encouragement and support in assisting us in completing this project within a very short time. We would also like to thank our Head of Department, **Dr. Ranjit KN**, for creating an atmosphere that promotes innovation and practical implementation of our academic curriculum. Lastly, we thank our Principal, **Dr. Y T Krishne Gowda**, for giving us a golden chance to undertake project on the subject of 'AI Powered Drug Discovery', and for his constant support in our research learning activity.

REFERENCES

[1]. Gupta, A., Wang, Y., & Zhao, J. (2021). Generative models for drug discovery: Recent advances and challenges. *Journal of Chemical Information and Modeling*, 61(3), 1239–1253.

[2]. Ahmed, Z., Zeeshan, S., Mendhe, D., & Dong, X. (2020). Human gene and disease associations for clinical-genomics and precision medicine research. *Clinical and Translational Medicine*, 10(5), e171.

DOI: 10.17148/IARJSET.2025.125203

[3]. Wang, Q., Li, J., & Wu, Z. (2020). Deep learning approaches for predicting drug-target interactions using graph-based models. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(3), 1180–1191.

[4]. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680.

[5]. Rehman, A., Khan, A., & Ahmad, M. (2023). AI- powered drug discovery: Current trends and future directions. *Computational and Structural Biotechnology Journal*, 21, 450–462.

[6]. Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., ... & Leach, A. R. (2017). The ChEMBL database in 2017. *Nucleic Acids Research*, 45(D1), D945–D954.

[7]. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J.,Marcu, A., Grant, J. R., ... & Wilson, M. (2018).Drug Bank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), D1074–D1082.

[8]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825– 2830.