Journal of Science, Technology and Social Transformation Volume 2, Issue 02, 2026

A Data-Driven Approach to Drug Discovery Analytics Using Generative Artificial Intelligence Techniques

Raghuvaran Kendyala

Department of Computer Science, University of Illinois at Springfield, USA

Abstract

Pharmaceutical industry is becoming more and more difficult in drug exploration and trial administration, as the expenses, long-term development, and more complex regulatory actions are rising. To respond to the mentioned challenges, the current research suggests a new model of drug discovery analytics that is constructed using a GAN, which is efficient in learning multidimensional patterns in molecules and enhancing predictive power. The data is preprocessed, features calculated by means of molecular descriptors, selected, and then the SMOTE is used to address unequal class distribution, and the model is then thoroughly trained and assessed. The suggested GAN model shows excellent performance, outperforming other ML and DL models with a (precision of 99.51), (recall of 99.99), (F1-score of 99.75) and (accuracy of 99.75). These findings imply that it has a high ability to generalize, less overfitting, and high robustness to classification tasks. On the whole, the framework provides an efficient, promising, and scalable idea in streamlining the drug discovery processes, facilitating better decisions, and speed-up innovation in pharmaceutical research and development.

Keywords: Drug Discovery, Generative AI, Molecular Design, ML, DL, Drug-Target Interaction. DOI:10.64235/jstst/221

Introduction

The combination of the data-driven approach and the creation of advanced AI has completely changed the process of drug discovery. The processes of conventional drug development are reported to be time-consuming, costly with high failure rates largely due to the complexity of the biological systems and the inability to predict. The massive collection of biomedical data, the genomic and proteomic data, to clinical trial reports in the past few years has created various new opportunities to use the computational methods to accelerate and simplify the drug discovery processes [1][2]safe, and effective. High-Performance Liquid Chromatography (HPLC. The past paradigm in this regard has been generative artificial intelligence, which has made it possible to create, predict, and optimize novel therapeutic candidates with previously unseen efficiency.

Machine-learning generative AI systems such as transformer-based models, VAEs and GANs can be trained and used to generate and train complex chemical and biological space. The models can be trained based on the underlying data distributions, and they can generate novel molecular structures which satisfy predefined pharmacological and physicochemical requirements [3] various high-quality protein crystal structures, and as a basis for computationally screening for compounds with improved inhibitory activity, bioavailability, and ADMETox properties. The ChEMBL and PubChem database contains experimental data from screening small molecules against SARS-CoV-2 3CLpro, which expands the opportunity to learn the pattern and design a computational model that can predict the potency of any drug compound against coronavirus before in-vitro and in-vivo testing. In this study, Utilizing several descriptors, we evaluated 27 machine learning classifiers. We also developed a neural network model that can correctly identify bioactive and inactive chemicals with 91% accuracy, on CheMBL data and 93% accuracy on combined data on both CheMBL and Pubchem. The F1-score for inactive and active compounds was 93% and 94%, respectively. SHAP (SHapley Additive exPlanations. Combining massive datasets with deep learning models, generative models are used to achieve de novo drug design, lead optimization, and predict drug-target interactions [4]. This transition to more data-driven generative frameworks and away from rule-based and heuristic data discovery approaches is a milestone breakthrough in computational drug discovery.

In addition, application of data analytics in generative AI processes can improve decision-making on various phases of drug development [5]. Advanced analytics methods, including feature engineering, dimensionality reduction,

Corresponding Author: Raghuvaran Kendyala Department of Computer Science, University of Illinois at Springfield, USA

How to cite this article: Kendyala R (2026). A Data-Driven Approach to Drug Discovery Analytics Using Generative Artificial Intelligence Techniques *Journal of Science, Technology and Social Transformation* 2(2), 1-8.

Source of support: Nil

Conflict of interest: None

and predictive modeling can be used to extract meaningful insights of heterogeneous data in the biomedical field [6] validations, and synthetic procedures are costly and time-consuming for drug discovery. Advancements in artificial intelligence (AI. These methods with generative models can aid in the discovery of new biomarkers, the optimization of compound libraries, and better toxicity and efficacy profiles prediction [7]. Therefore, the process of discovery is not only faster in data-driven generative AI systems, but increases its reliability and reproducibility.

Although such improvements have been made, there are still a number of issues with the application of generative AI in drug discovery analytics [8]. Data quality, model interpretability, scalability and regulatory compliance are issues that are to be handled with care to make sure that they are practical to use [9]. Moreover, domain knowledge combined with machine learning models is also a research problem that is yet to be resolved, and involves the interdisciplinary efforts of computational scientists, chemists, and clinicians. This paper is a detailed analysis of the subject of data-driven drug discovery analytics through generative AI technology, current trends, analytical processes and focuses on future studies and research that can reshape pharmaceutical innovation.

Motivation and Contribution of Paper

The conventional drug discovery method is expensive, long and in most cases, it is defined by high failure rates since the biological systems are complex and there is the inability to accurately predict how the molecules behave. Even though recent developments in ML and DL have enhanced predictive performance, most of the existing models are unable to deal with high-dimensional inputs, deal with imbalanced classes and effectively capture complex relationships among molecules. Also, overfitting, poor generalization, and poor accuracy are some of the problems that inhibit their practical applicability. Thus, there is a demand to have a robust, data-driven method that will be efficient to model complicated patterns, produce high-quality synthetic data, and increase the accuracy of predictions, which is why the application of GANs is the best way to develop better drug discovery analytics. The main findings of this study are as follows:

- The new data-driven model of drug discovery is introduced that involves incorporating generative artificial intelligence methods to improve predictive modeling and molecular analysis.
- The complex representations of the molecules and also enhance the classification performance, a Generative Adversarial Network (GAN)-based architecture is created.
- A preprocessing pipeline thatincludes datatransformation, extraction of molecular descriptors, feature selection and class imbalance mitigation operations are planned to guarantee the quality of data and the efficiency of the model.
- Standard performance is used to demonstrate the strength and efficacy of the suggested model. measures with the help of an extensive experimental assessment.

Novelty and Justification

The novelty of the work is in the fact that a Generative Adversarial Network (GAN) has been incorporated into a full-scale data-driven platform of drug discovery analytics, which allows learning complex molecular patterns and predicting better. However, in comparison to custom ML and isolated DL methods, the suggested model will use a full preprocessing pipeline, such as feature selection and handling imbalances between the classes to strengthen the model's resilience and the quality of the data. This study is justified by the necessity to have proper, scalable and stable computational procedures to counter the shortcomings of traditional drug discovery procedures, including high cost, time, and low success rates. The offered method is a worthy addition to the pharmaceutical research of the present days, as it enables the better prediction and exploration of the molecular space efficiently, which becomes possible with the support of generative AI.

Outline of the Paper

The rest of the document is designed as follows: Part II offers the literature review and outlines the methods that are in use in drug discovery. Section III outlines the suggested methodology, such as the preprocessing of data, feature selection, the aspect of class imbalance, and the GAN-based model. Section IV gives the presentation of the performance evaluation and experimental results. Lastly, Section V is the conclusion of the paper, and seems to be the directions to the further research.

Literature Review

Early implementations of generative AI in drug discovery primarily used rule-based and supervised learning approaches as summarized in Table I.

G. Sumathi et al. (2024) explored potential way to find new treatments and solve medical problems that have so far gone unsolved is to include GANs into molecular design processes. Results from Drug Design Data Resource, Kaggle datasets shows that the molecular formula with id Activity-Based Learning (ABL) minimum value starts with 0.055 and maximum of 10, from the sample of 10 compounds the minimum affinity value is 0.075 and maximum of 3.42. In another data surveyed the drug discovery values show the minimum of the molecular weight is 356.3 and maximum is 471.5 [10].

Similarly, R. Madunuri et al. (2024) utilized general approach entails vast data gathering, preprocessing, and efficient feature extraction techniques and the application of many ML models, including RF, CNN, SVM, and RNN. The experimental results presented also establish CNN with an accuracy of 92%, the effectiveness of computational methods and blowing superiority in the experimental verifications [11].

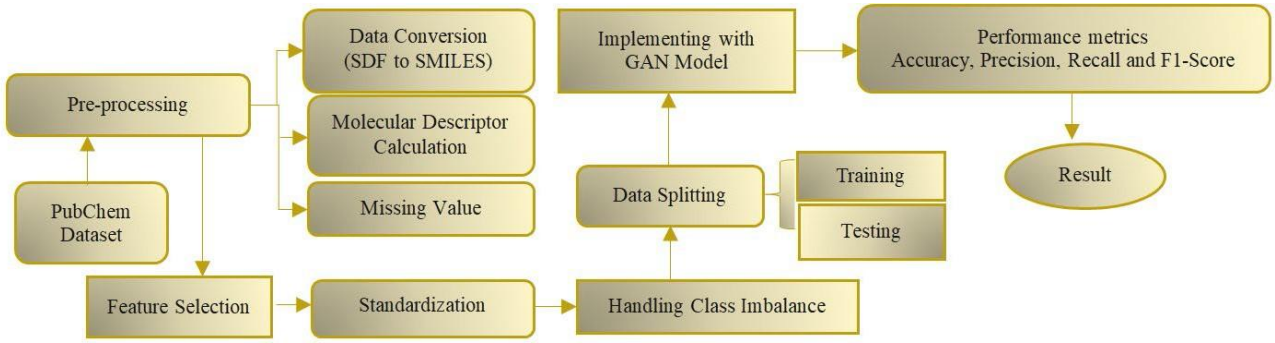

Figure 1: Flowchart of Drug Discovery using GenAI model

Furthermore, G. Hanson et al (2024) used the Logistic Regression technique to create a solid three-dataset splitting strategy that produced a 94% accuracy rate. Additional analysis of 2683 new substances from the EANPDB and ZINC databases was made possible by the model's successful prediction of 18 known DENV inhibitors, of which 11 were shown to be active [1]. In another study, S. R. Burri et al. (2023) proposed a identifies the enforcement issues in this industry and provides remedies. Another area of research is an automatic fault diagnosis system for natural leather based on Generative Adversarial Networks (GANs). Despite its shortcomings and biases, the results demonstrate that the recommended approach is accurate and cost-effective. AI has transformed precision medicine, drug discovery, and medical diagnostics, giving this work significance [12].

Additionally, M. A. Rezaei et al. (2022) proposed a data-driven framework named DeepAtom to accurately predict the affinity for protein-ligand binding. DeepAtom might use the Using Binding-related atomic interaction patterns can be automatically retrieved from the voxelized complex structure using a 3D Convolutional Neural Network (3D-CNN) architecture. Even with the small amount of training data available, their lightweight model architecture significantly increases the model's representational capacity when compared to existing CNN-based methods. They also develop and recommend a new benchmark dataset to further improve the model's performance. Using the new dataset as training input, DeepAtom achieves Pearson's R=0.83 and RMSE=1.23 pK units on the PDBbind v.2016 core set [13].

Lastly, M. Xu et al. (2021) created a generative model for de novo small molecules called DeepGAN, which uses the Generative Adversarial Network algorithm as its foundation. The model is trained to maximize adversarial loss and rewards in certain domains using a strategy gradient. Thus, DeepGAN performs better than ORGAN and its offshoots, OR(W)GAN and Naive RL, which have previously undergone extensive testing. According to the experiments, their approach can produce compounds that preserve chemical diversity, boost validity, and demonstrate progress in the targeted metrics [14].

Although AI-based drug discovery has improved, there exists a limitation in the existing techniques to manage high-dimensional data and identify intricate chemical patterns. Both DL and conventional ML models have reasonable accuracy, and they do not generate new molecular structures. Despite the potential of GAN-based approaches, the problems of instability during training, weak generalization, and poor management of the class imbalance are still present. Consequently, an integrated, data-driven generative framework is required to make sure that, in drug discovery applications, robust learning is guaranteed, data is represented in a balanced way, and the predictive performance is enhanced.

METHODOLOGY

The proposed drug discovery approach involving Generative Artificial Intelligence has a well-organized pipeline as depicted in Fig. 1, where data is first collected and then preprocessed via data conversion to SMILES, and molecular descriptors computation as well as managing missing values to ensure that the data quality is appropriate. The feature selection is then done to select relevant features and then class imbalance handling is done through techniques like SMOTE in order to have a balanced dataset. The processed Training and testing sets of data are separated and a Generative Adversarial Network (GAN) model is applied, with the generator and discriminator being trained alternately to extract learned complex patterns. Lastly, the model is assessed through to demonstrate the effectiveness of the proposed model, measures of performance such as recall, precision, accuracy, and F1-score are compared with the outcomes.

This section explains the following phases of the flowchart in a nutshell

Data Collection and Analysis

PubChem is a publicly accessible chemical database that integrates information from multiple sources, providing data on chemical compounds and their bioactivities. It includes both small and large molecules and serves as an important resource for drug discovery and virtual screening. It also facilitates computational research through provision of

Table 1: Summary of Related Works in AI-Driven Drug Discovery

Author	*Methods and Techniques*	*Data*	*Key Findings*	*Limitations/ Challenges*	*Future Implications*
G. Sumathi et al. (2024)	GAN for molecular design	Drug Design Data Resource, Kaggle	Molecular activity ranges from 0.055–10; affinity from 0.075–3.42; molecular weight from 356.3–471.5	Limited sample size and lack of real-time validation	Improve dataset diversity and validate with experimental studies
R. Madunuri et al. (2024)	SVM, RF, CNN, RNN	Preprocessed drug datasets	CNN achieved 92% accuracy with better efficiency and prediction	Computational complexity and dependency on feature engineering	Optimize models and explore hybrid deep learning techniques
G. Hanson et al. (2024)	Logistic Regression with dataset splitting	ZINC, EANPDB	Achieved 94% accuracy; identified active inhibitors with high binding affinity	Limited model complexity and scalability issues	Apply deep learning for improved prediction and scalability
S. R. Burri et al. (2023)	GAN-based detection system	Industrial/ medical datasets	High accuracy and cost-effective solution using GAN	Bias and generalization limitations	Improve robustness and reduce bias in GAN models
M. A. Rezaei et al. (2022)	3D-CNN (DeepAtom)	PDBbind v.2016, Astex	RMSE = 1.23 and Pearson's R = 0.83, outperforming baseline models	Requires large computational resources	Develop lightweight and efficient architectures

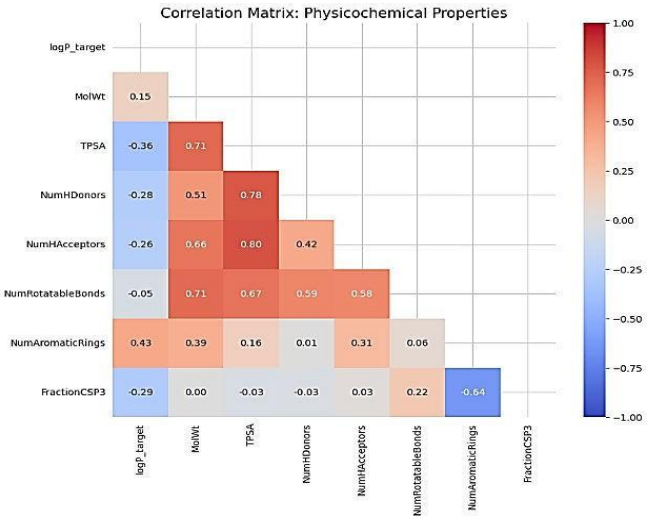

Figure 2: Correlation matrix of Physical-chemical description.

systematic and standardized information that can be used in machine learning processes.

The correlation matrix of physicochemical descriptors is shown in the Fig. 2, where the strengths and direction of relationships between the features is shown. TPSA and NumHAcceptors (0.80), TPSA and NumHDonors (0.78), and MolWt have significant positive correlations with TPSA (0.71). NumRotatableBonds and NumHAcceptors are also moderate (0.67 and 0.58) correlations with TPSA. Target (logP target) is weakly correlated with most of the features, including MolWt (0.15) and TPSA (-0.36). FractionCSP3 and NumAromaticRings are found to have a notable negative correlation (-0.64). On the whole, the matrix identifies significant relationships and multicollinearity between the descriptors.

Data Preprocessing

Drug discovery the preprocessing on PubChem data consists of SDF-to-SMILES, PaDEL-Descriptor feature descriptors, A mean to replace missing values, feature selection based on variance, data standardization, and the welfare of class

imbalance without oversampling. This process entails the following steps that are discussed below:

Data Conversion (SDF to SMILES)

Chemical structures are converted into SMILES format, a simplified textual representation that is easier to process and use in machine learning applications.

Molecular Descriptor Calculation

Chemical structures are transformed into numerical features representing physicochemical properties using tools like PaDEL-Descriptor, enabling machine learning modelling.

Missing Value Imputation

The mean imputation is utilized to fill in the dataset's missing values with the average of the missing features so that no feature is left out in the training of the model.

Feature Selection (Variance Filtering)

The feature selection process involves the identification and retention of the most relevant features and the removal of redundant features and even unnecessary features [15]. Variance filtering eliminates features that vary not significantly across samples because they do not play a significant role in classification between classes. The step enhances efficiency and performance of the models.

Feature Scaling (Standardization)

The feature scaling normalizes the range of features enabling them to have an equal contribution in the model. With the help of such tools as the StandardScaler of scikit-learn, data are rescaled to a means zero and the standard deviation equal to one. The majority of machine learning algorithms' performance and stability will be greatly improved by this step.

Handling Class Imbalance

Class imbalance in the classification problem is a situation where one of the classes has very many representatives and the other has only a few leading to skewed model results towards the prevalent class. In order to deal with this problem, to balance the data, artificial samples of the minority category are created applying the SMOTE, as opposed to just replicating the examples that already exist.

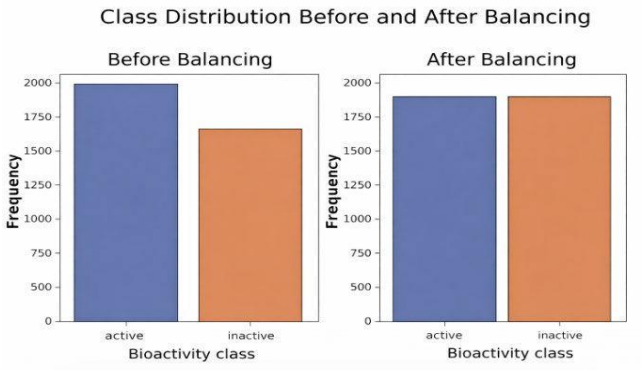

Figure 3: Class distribution before and after balancing with SMOTE.

Fig. 3 shows the distribution of bioactivity classes both prior to and following the application of The method known as Synthetic Minority Over-sampling (SMOTE). First, the dataset has an uneven distribution of classes with the active class having more samples than the inactive class. Following SMOTE, artificial samples of the minority group are created leading to an equal representation of active and inactive groups.

Data Separation

Additionally, the dataset has been separated into training (80%) and testing (20%). The model is built and trained using the training data and the final performance is assessed with the testing data.

Proposed GAN Model

The given model is grounded in the concept of a Drug discovery analytics is the usage of the Generative Adversarial Network (GAN) system in order to generate and classify the patterns of molecular data [16]. The two main components of the GAN are the discriminator that is trained to differentiate between generated and authentic samples and the generator, which is trained to produce valid molecular representations. The adversarial training also enhances the model's accuracy. To keep the model alive, the Adam optimizer is used with a beta of 0.5 and a learning rate of 0.0002. The best ratio between learning and generalization is a batch size of 64 and 20 training epochs. Besides, the network uses sigmoid on the output layer and ReLU on the hidden layers where the loss function is binary cross-entropy to direct optimization. Equation (1) derives it:

$$L = -[y log(y^{`}) + (1 - y) log(1 - y^{`})] \quad (1)$$

Performance Metrics

Some popular classifications metrics including accuracy, precision, recall, and F1-score are used to evaluate the model's effectiveness. The TP, TN, FP, and FN components of the confusion matrix serve as the foundation for these metrics:

True Positive (TP)

Accurately anticipated positive instances (Active substances correctly classified as Active).

True Negative (TN)

This is the quantity of negative instances that were accurately anticipated to be inactive (i.e., inactive substances).

False Positive (FP)

Missing positive examples (Inactive compounds classified as Active).

False Negative (FN)

Actives labelled as inactives.

The proportion of accurate forecasts among all forecasts is called accuracy, and it is the general accuracy of the model [16]. The fraction of the positive instances that will be correctly predicted out of all the positives predicted, is referred to as precision. Recall (Sensitivity) This is a metric as to the number of instances of positive instances that exist which the model recognizes correctly. The F1-score is a dependable and solid assessment of the model's functionality in drug discovery analytics since it is the harmonic mean of precision and recall. Metrices derived in Equations (2)-(5):

$$Accuracy = \frac{TP+TN}{TP+Fp+TN+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$F1-Score = 2\frac{(Precision*Recall)}{Precision+Reall} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Result Analysis and Discussion

A high-performance computing environment using an Intel Core i7 processor and an NVIDIA GeForce RTX 2080 Ti GPU 8700K processor running at 3.70 GHz to up deep learning calculations, and 32 GB of RAM to effectively manage big datasets is used for the experimental setup. The primary programming language utilized to implement and carry out the tests, and the software setup is based on the Windows 10 operating system. In Table II the proposed Generative Adversarial Network (GAN) model on the PubChem dataset proves to have an outstanding predictive power in all the metrics of evaluations. The model's accuracy of 99.75 indicates that its performance is quite dependable. The percentage accuracy (99.51) suggests a low-volume of false positive and recall (99.99) indicates the model is a working tool which detects almost all the relevant cases correctly. Moreover, 99.75% F1-score demonstrates that the suggested method exhibits the robustness and efficacy of the suggested methodology in drug discovery analytics and strikes a good balance between recall and precision.

Fig. 4 shows accuracy of the proposed Generative Adversarial Network (GAN) model in terms of training and testing accuracy through various epochs. As it is possible to see, the accuracy of the training rises gradually at the beginning with around 91% and then rises to over 99%, which demonstrates the successful acquisition of the model. At the same time, the accuracy of testing is quite high with the point of about 98% and plateauing at 99.5% and this is indicative of good generalization. The low separation among training and testing shows accuracy that the model performs well with the unknown data and is not overfitting.

Fig. 5 illustrates the testing and training loss curve of the proposed Generative Adversarial Network (GAN) model with various epochs. The training loss depicts a sharp decrease between the range of 0.21 to less than 0.04, which demonstrates that the model minimizes error well in the process of learning. On the same note, the testing loss diminishes at a high rate during the first epochs and levels around 0.02, indicating a consistent performance on the

Table 2: Accuracy Plot of GAN model.

Metrics	*Accuracy*	*Precision*	*Recall*	*F1-score*
GAN	99.75	99.51	99.99	99.75

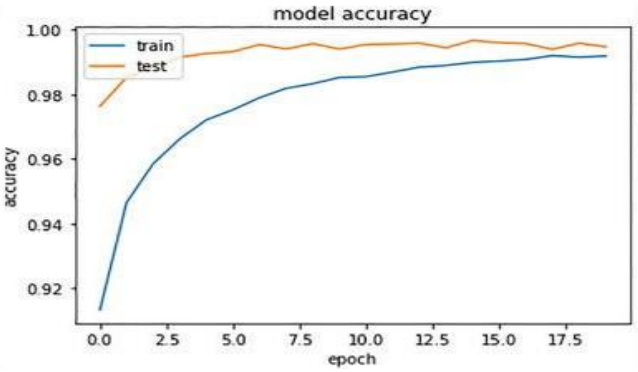

Figure 4: utcomes of the Suggested Model's PubChem Dataset

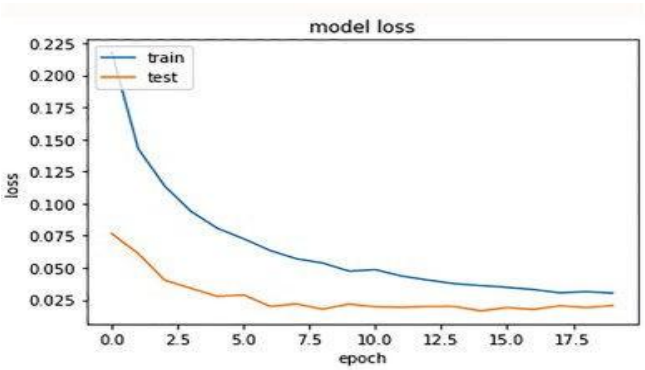

Figure 5: Loss Plot of GAN model.

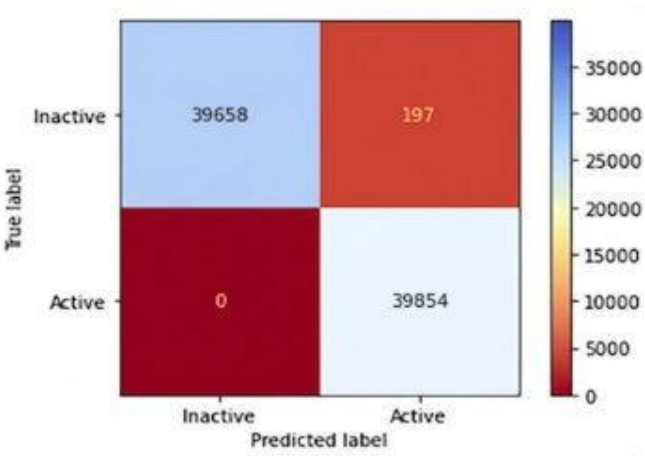

Figure 6: GAN model Confusion Matrix.

Table 3: Ml And Dl Model Comparison For Drug Discovery

Measures	*Accuracy*	*Precision*	*Recall*	*F1-score*
GAN	99.75	99.51	99.99	99.75
LGBM[17]	76.4	-	76.3	-
MLP[17]	77.8	-	78	-
CNN[19]	8.0	51.4	86.2	64.44
DeepDTI[19]	70	37.8	89.7	53.3

unseen information.

The suggested GAN model's confusion matrix of the PubChem dataset is shown in Fig. 6. The model has 39,658 inactive and 39,854 active compounds correctly predicted which illustrates high accuracy of the predictions in both classes. There are only 197 misclassified inactive samples as active and no false negatives are being predicted as such, zero. Here in lies the extreme sensitivity and capability of the model to accurately find each active compound which is paramount in the drug discovery settings.

Comparison and Discussion

A comparison of various ML and DL drug discovery, in Table III displays models according to precision, accuracy, recall, and F1-score. The recommended GAN model maintains a consistent and high F1-score, recall and precision with the highest accuracy of 99.75%, which result in a high degree of reliability. Classical models such as Random Forest are effective with more than 90 percent accuracy, and LGBM and MLP have an intermediate performance of 76-78 %. However, neural networks like CNN and DeepDTI do not show consistent results with unequal metrics. In general, the GAN model performs better than the rest as depicted in Table III to illustrate its power in drug discovery endeavors.

The model proposed applies A generator and discriminator make up a GAN which is trained in a manner of learning complex patterns of data via adversarial training. This will allow high accuracy in the model, as well as stability in the performance, as it is able to establish the complex relationships in data. Its main strengths are that it has higher predictive power, good generalization, lesser overfitting, and capable of generating realistic artificial data, which makes it very useful in drug discovery research.

Limitations

Although the proposed model has a good performance, it has limitations. The GAN system is prone to unstable training and needs to be carefully set up in order to produce the best results. Also, the model is very dependent on the input data quality and representation, and it may affect generalization in case of noisy or biased data. The adversarial learning process implies a fairly high computational complexity and a training time. Moreover, it is not easily interpretable and therefore hard to follow the decision-making process and there is no real-world experimental validation that is needed by the model to be practical in drug discovery applications.

Conclusion and Future Direction

Generative AI is transforming drug discovery from a linear, high-risk process to an adaptive, intelligent system. By accelerating molecular design, optimizing lead compounds, and personalizing medicine, these technologies hold the promise to redefine pharmaceutical innovation. This paper is a data-based drug discovery framework based on Generative Artificial Intelligence, specifically, it builds a Generative Adversarial Network (GAN) to predict complicated molecular structures and improve predictive accuracy. The proposed solution combines successful preprocessing, feature selection, a class imbalance (SMOTE) and high-quality model training to obtain great precision and dependability of the findings. The experiment findings show demonstrates in terms of accuracy, precision, recall, and F1-score, the GAN model outperforms deep learning and traditional machine learning models demonstrate how much more effective it is. The high level of consistency in evaluation parameters, the ability to generalize and the low overfitting rates point to the high level of efficiency of the proposed methodology in enhancing and speeding up drug discovery procedures.

Further research can be conducted on how to make models more interpretable and also incorporate domain knowledge to make decisions more effective in drug discovery. Moreover, it is possible to consider more sophisticated generative models like diffusion models and transformer-based models to enhance the performance of molecular generation and prediction. Increasing the framework by applying real-time experimental validation and bigger and more diversified datasets will also contribute to making the proposed approach more practical and scalable.

References

[1] G. Hanson *et al.*, "Machine learning and molecular docking prediction of potential inhibitors against dengue virus," *Front. Chem.*, vol. 12, Dec. 2024, doi: 10.3389/fchem.2024.1510029.

[2] R. Snehamrutha, "Integration of Analytical Techniques (HPLC & FTIR) for Enhanced Quality Assurance in Drug Formulation," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 1, p. 953, Feb. 2026, doi: 10.48175/IJARSCT-12750C.

[3] F. Bin Ashraf, S. Akter, S. H. Mumu, M. U. Islam, and J. Uddin, "Bio-

activity prediction of drug candidate compounds targeting SARS-Cov-2 using machine learning approaches," *PLoS One*, vol. 18, no. 9, p. e0288053, Sep. 2023, doi: 10.1371/journal.pone.0288053.

[4] V. Patel and M. Shah, "Artificial intelligence and machine learning in drug discovery and development," *Intell. Med.*, vol. 2, no. 3, pp. 134–140, Aug. 2022, doi: 10.1016/j.imed.2021.10.001.

[5] A. Daina and V. Zoete, "Testing the predictive power of reverse screening to infer drug targets, with the help of machine learning," *Commun. Chem.*, vol. 7, no. 1, p. 105, May 2024, doi: 10.1038/s42004-024-01179-2.

[6] W. Chen, X. Liu, S. Zhang, and S. Chen, "Artificial intelligence for drug discovery: Resources, methods, and applications," *Mol. Ther. - Nucleic Acids*, vol. 31, pp. 691–702, Mar. 2023, doi: 10.1016/j.omtn.2023.02.019.

[7] H. Gajjar, U. Kora, and V. Gondaliya, "Telemedicine with Automated ECG Analysis," *Int. J. Eng. Res. Technol.*, vol. 2, no. 10, pp. 2253–2255, 2013.

[8] D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, and R. K. Tekade, "Artificial intelligence in drug discovery and development," *Drug Discov. Today*, vol. 26, no. 1, pp. 80–93, Jan. 2021, doi: 10.1016/j.drudis.2020.10.010.

[9] M. K. Yadav *et al.*, "Unleashing the future: The revolutionary role of machine learning and artificial intelligence in drug discovery," *Eur. J. Pharmacol.*, vol. 985, p. 177103, Dec. 2024, doi: 10.1016/j.ejphar.2024.177103.

[10] G. Sumathi, P. S. Juliet, M. S. Banu, S. D. A. Sujitha, Y. S, and S. Sujatha, "Molecular Design for Drug Discovery with the Use of Generative Adversarial Networks," in *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, IEEE, Aug. 2024, pp. 885–890. doi: 10.1109/ICESC60852.2024.10689846.

[11] R. Madunuri, S. M. Yellepeddi, C. S. Ravi, S. Chitta, V. S. M. Bonam, and V. K. R. Vangoor, "AI-Enhanced Drug Discovery Accelerating the Identification of Potential Therapeutic Compounds," in *2024 Asian Conference on Intelligent Technologies (ACOIT)*, IEEE, Sep. 2024, pp. 1–8. doi: 10.1109/ACOIT62457.2024.10939855.

[12] Nadella, N. V. M. (2025a). Evolution of fault management in telecommunications: From reactive response to AI-driven predictive analytics. World Journal of Advanced Engineering Technology and Sciences, 15(2), 1338– 1344. https://doi.org/10.30574/wjaets.2025.15.2.0671

[13] S. R. Burri, M. Y. Diallo, L. Sharma, and V. Dutt, "AI-Driven Drug Discovery: Unravelling the Potential of Generative Adversarial Networks (GANs) in Pharmaceutical Research," in *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, IEEE, Nov. 2023, pp. 817–822. doi: 10.1109/ICTACS59847.2023.10390116.

[14] Nadella, V. M. (2025). The backbone of modern society: How telecom technologies drive societal transformation. *World Journal of Advanced Research and Reviews*, *26*(2), 1425-1434.

[15] M. A. Rezaei, Y. Li, D. Wu, X. Li, and C. Li, "Deep Learning in Drug Design: Protein-Ligand Binding Affinity Prediction," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 19, no. 1, pp. 407–417, Jan. 2022, doi: 10.1109/TCBB.2020.3046945.

[16] M. Xu, J. Cheng, Y. Liu, and W. Huang, "DeepGAN: Generating Molecule for Drug Discovery Based on Generative Adversarial Network," in *2021 IEEE Symposium on Computers and Communications (ISCC)*, IEEE, Sep. 2021, pp. 1–6. doi: 10.1109/ISCC53001.2021.9631396.

[17] M. Kokudeva, M. Vichev, E. Naseva, D. G. Miteva, and T. Velikova, "Artificial intelligence as a tool in drug discovery and development," *World J. Exp. Med.*, vol. 14, no. 3, Sep. 2024, doi: 10.5493/wjem.v14.i3.96042.

[18] Nadella, V. M. (2024). AI-Native 6G Network Management. *American International Journal of Computer Science and Technology*, *6*(1), 23-37.

[19] A. Majeed and S. O. Hwang, "Data-Driven Analytics Leveraging Artificial Intelligence in the Era of COVID-19: An Insightful Review of Recent Developments," *Symmetry (Basel).*, vol. 14, no. 1, p. 16, Dec. 2021, doi: 10.3390/sym14010016.

[20] Mazumder, P. T. (2025). Blockchain in trade finance: reducing fraud and improving efficiency through digital ledger technology. Digital Finance, 7(4), 1043-1063.

[20] T. Yu, L. C. Chong, C. Nantasenamat, N. Anuwongcharoen, and T. Piacham, "Machine Learning Approaches To Study the Structure-Activity Relationships of Lpxc Inhibitors," *EXCLI J.*, vol. 22, pp. 975–991, 2023, doi: 10.17179/excli2023-6356.

[21] Nadella, V. M. (2025). AI/ML-Driven Service Assurance: 2024 breakthroughs transforming telecom operations. *Journal of Computer Science and Technology Studies*, *7*(4), 01-07.

[22] E. Toni, H. Ayatollahi, R. Abbaszadeh, and A. Fotuhi Siahpirani, "Machine Learning Techniques for Predicting Drug-Related Side Effects: A Scoping Review," *Pharmaceuticals*, vol. 17, no. 6, p. 795, Jun. 2024, doi: 10.3390/ph17060795.

[23] Y. Chen, X. Liang, W. Du, Y. Liang, G. Wong, and L. Chen, "Drug–Target Interaction Prediction Based on an Interactive Inference Network," *Int. J. Mol. Sci.*, vol. 25, no. 14, p. 7753, Jul. 2024, doi: 10.3390/ijms25147753.

[24] Sannapureddy, R., Nadella, V. M., & Nelavelli, S. (2024). Edge–Cloud Continuums for Latency-Sensitive Tasks. International Journal of AI, BigData, Computational and Management Studies, 5(4), 189-201.

[25] Madhav Nadella, V. (2025). under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0) International License Real-Time Monitoring Systems (5G/6G & Beyond): A Technical Review. Sarcouncil Journal of Engineering and Computer Sciences, 4(08).